

**Molecular simulations on proteins of
biomedical interest:**
A. Ligand–protein hydration
B. Cytochrome P450 2D6 and 2C9
C. Myelin Associated Glycoprotein (MAG)

Inauguraldissertation

zur

Erlangung der Würde eines Doktors der Philosophie
vorgelegt der
Philosophisch-Naturwissenschaftlichen Fakultät der
Universität Basel

von

Gianluca Rossato
aus Venedig, Italien

Basel, 2011

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät

Auf Antrag von:

Prof. Dr. A. Vedani, Institut für Molekulare Pharmazie, Universität Basel

Prof. Dr. S. Moro, Department of Pharmaceutical Sciences, University of Padova

Basel, den 21. Juni 2011

Prof. Dr. Martin Spiess

Dekan

Non pretendiamo che le cose cambino se continuiamo a farle allo stesso modo.

La crisi è la miglior cosa che possa accadere a persone e interi paesi perchè è proprio la crisi a portare il progresso.

La creatività nasce dall'ansia, come il giorno nasce dalla notte oscura. E' nella crisi che nasce l'inventiva, le scoperte e le grandi strategie.

Chi supera la crisi supera se stesso senza essere superato. Chi attribuisce le sue sconfitte alla crisi, violenta il proprio talento e rispetta più i problemi che le soluzioni.

La vera crisi è la crisi dell'incompetenza.

Lo sbaglio delle persone è la pigrizia nel trovare soluzioni.

Senza crisi non ci sono sfide, senza sfide la vita è una routine, una lenta agonia.

Senza crisi non ci sono meriti. E' nella crisi che il meglio di ognuno di noi affiora, perchè senza crisi qualsiasi vento è una carezza.

Parlare di crisi è creare movimento; adagiarsi su di essa vuol dire esaltare il conformismo.

Invece di questo, lavoriamo duro!

L'unica crisi minacciosa è la tragedia di non voler lottare per superarla.

Albert Einstein (?)

Let's not pretend that things will change if we keep doing the same things.

A crisis can be a real blessing to any person, to any nation, because the crisis brings progress.

Creativity is born from anguish, just like the day is born from the dark night. It is in crisis that inventive is born, as well as discoveries and big strategies.

Who overcomes crisis, overcomes himself, without getting overcome. Who blames his failures to a crisis, neglects his own talent, and is more respectful to problems than to solutions.

Incompetence is the true crisis.

The greatest inconvenience of people and nations is the laziness with which they attempt to find the solutions to their problems.

There's no challenge without a crisis. Without challenges, life becomes a routine, a slow agony.

There is no merit without crisis. It is in the crisis where we can show the very best in us. Without a crisis, any wind becomes a tender touch.

To speak about a crisis is to promote it, Not to speak about it is to exalt conformism.

Let us work hard instead!

Let us stop, once and for all, the menacing crisis that represents the tragedy of not being willing to overcome it.

Albert Einstein (?)

Acknowledgements

First, I would like to thank Prof. Dr. Angelo Vedani, for the possibility he gave me to work in his research group, for his guidance in the project, and for the freedom he left me in the development of my ideas. I personally think that the term “father professor” perfectly fits to the role he had in this period of my professional life, conjugating both its importance and the ambivalent moments of agreement and disagreement, which naturally characterize this particular relationship.

A special “thank you” goes to Prof. Dr. Beat Ernst, for the stimulating environment, for the enthusiasm, for the help in improving my writing skills, and for the many fruitful collaborations we had during my PhD studies.

One of the most potent thanks goes to Dr. Martin Smiesko, *in primis* Friend, which in these three years shared with me moments of happiness, discouragement, and of garden digging. Secondly, I'm grateful for his professional advices, for his secure knowledge, and for his valuable teaching skills that he applied to me both in the modeling and in the programming field.

A sincere thanks goes to Prof. Dr. Stefano Moro, for accepting to be my co-referee, for his trust that allowed me to initiate my career in drug discovery, and for his mentorship in my professional and personal life.

I'm grateful to Prof. Emeritus Max Dobler for the valuable software he supplied me during my studies and for his continuous support provided on BioX, and BiografX. Prof. Dr. Mark S. Johnson and Dr. Henri Xhaard for being part of my Finnish experience and for seeing in me a potential good researcher.

Alla mia famiglia: a mia madre Lorenza, mio padre Enzo e mia sorella Martina, che hanno sempre assecondato le mie scelte, comprendendole e dandogli valore. Grazie per avermi sempre aiutato in tutti gli aspetti e per essermi sempre stati vicini. Ai miei nonni Nina, Franco e Liliana per avermi ricordato e per ricordarmi sempre che cosa sia importante.

Thank you to the modeling group: to Morena.. you are of the most important person of this experience, which helped me in any sort of thing. Without you my integration in Basel would have been much more difficult. To Rania, which despite her silence shared so many moments in the lab. To Sameh and Adam that brought a wind of change in 3026. To the group of the IMP, for the great atmosphere and friendship demonstrated in different moments. Hoping to not

forget anybody thanks to Gabi, Christa, Meike, Kathi, Jaqueline, Daniela, Katrin, Stefanie, Arjan, Jonas, Matthias, Roland, Florian, Kiki, Brian, Oliver, Said, Bea, Simon, Mirko, Matthias G. Wojciech, Lijuan, Katja, Xiaohua.

Thank you to all the people which made so nice the time outside of the lab: Daniel for the uncountable evenings in Cargo, Ginger for the thousand of sporty moments spent together and for the many MTB tours done and to come. Thanks to Thomas, Daniela B., Morena & Matthias W, Molly and Lucia for the nice moments spent together and to Annalisa, Maria, Marta and Elena for the company in the never-ending queues out of the Gotthard tunnel.

To Ela for coming into my life and for making it more colorful and valuable.

Thank you to the members of my WG: Stefan, Aida, Joelle, Till, Richi, Katrin, and Lena. To my Italian friends that made me always feel home and welcome every time I was going back to Italy. Thank you Massimo, Michela, Filippo and Lisa for the many erasmus reunions. Thank you Marco, Cecilia and Fra, to always make me feel good.

The last special thanks goes to Basel for the “love and hate” relationship we had so far and to the Rhein being my favorite anti-stress place.

The Swiss National Science Foundation is gratefully acknowledged for the financial support to the project.

Abstract

TOPIC 1: Water molecules mediating polar interactions in ligand–protein complexes contribute to both binding affinity and specificity. To account for such water molecules in computer-aided drug discovery, we performed an extensive search in the Cambridge Structural Database (CSD) to identify the geometrical criteria defining interactions of water molecules with ligand and protein. In addition, *ab initio* calculations were used to derive the propensity of ligand hydration. Based on these information we developed an algorithm (AcquaAlta) to reproduce water molecules bridging polar interactions between ligand and protein moieties. This approach was validated using 20 crystal structures and yielded a match of 76% between experimental and calculated water positions. The solvation algorithm was then applied to the docking of oligopeptides to the periplasmic oligopeptide binding protein A (OppA), supported by a pharmacophore-based alignment tool.

TOPIC 2: Drug metabolism, toxicity, and interaction profile are major issues in the drug discovery and lead optimization processes. The Cytochromes P450 (CYPs) 2D6 and 2C9 are enzymes involved in the oxidative metabolism of a majority of the marketed drugs. By identifying the binding mode using pharmacophore pre-alignment and automated flexible docking, and quantifying the binding affinity by multi-dimensional QSAR, we validated a model family of 56 compounds (46 training, 10 test) and 85 (68 training, 17 test) for CYP2D6 and CYP2C9, respectively. The correlation with the experimental data (cross-validated $r^2 = 0.811$ for CYP2D6 and 0.687 for CYP2C9) suggests that our approach is suited for predicting the binding affinity of compounds towards the CYP2D6 and CYP2C9. The models were challenged by Y-scrambling, and by testing an external dataset of binding compounds (15 compounds for CYP2D6 and 40 for CYP2C9) and not binding compounds (64 compounds for CYP2D6 and 56 for CYP2C9).

TOPIC 3: After injury, neurites from mammalian adult central nervous systems are inhibited to regenerate by inhibitory proteins such as the myelin-associated glycoprotein (MAG). The block of MAG with potent glycomimetic antagonists could be a fruitful approach to enhance axon regeneration. Libraries of MAG antagonists were derived and synthesized starting from the (general) sialic acid moiety. The binding data were rationalized by docking studies, molecular dynamics simulations and free energy perturbations on a homology model of MAG. The pharmacokinetic profile (i.e. stability in cerebrospinal fluid, logD, and blood-brain barrier permeation) of these compounds has been thoroughly investigated to evaluate the drug-likeness of the identified antagonists.

List of Publications

Publications included in the thesis:

This thesis is based on the listed publications, which are referred to in the text with the following numbers:

Publication 1: G.Rossato, B.Ernst, M. Smiesko, M.Spreafico, A.Vedani. Probing small-molecule binding to Cytochrome P450 2D6 and 2C9: An in silico protocol for generating toxicity alerts. *ChemMedChem*. **2010**, 5, 2088-2011

Publication 2: O. Schwardt, H. Koliwer-Brandl, R. Zimmerli, S. Mesch, G. Rossato, M. Spreafico, A.Vedani, S. Kelm, B. Ernst. Design, synthesis, biological evaluation, and modeling of a non-carbohydrate antagonist of the myelin-associated glycoprotein. *Bioorg. Med. Chem.* **2010**, 18, 7239-7251

Publication 3: S. Mesch, D. Moser, D. S. Strasser, A. Kelm, B. Cutting, G. Rossato, A.Vedani, H. Koliwer-Brandl, M. Wittwer, S. Rabbani, O. Schwardt, S. Kelm, B. Ernst, Low Molecular Weight Antagonists of the Myelin-Associated Glycoprotein: Synthesis, Docking, and Biological Evaluation. *J. Med. Chem.* **2010**, 53(4),1597-615

Publication 4: G.Rossato, B. Ernst, A. Vedani, M. Smiesko, AcquaAlta: A Directional Approach to the Solvation of Ligand–Protein Complexes. *J. Chem. Inf. Mod.* **2011**, *in press*

Note that “Publication 4” has been added as “Appendix 5” after the PhD exam, as at the moment of the thesis submission the manuscript was accepted but still not “in press” in the current format.

The author's contribution

Publication 1: G.Rossato et al. "Probing small-molecule binding to Cytochrome P450 2D6 and 2C9: An in silico protocol for generating toxicity alerts."

This project both in term of planning, simulations, calculations, and analysis was completely carried out by me.

Publication 2: O.Schwardt et al. "Design, synthesis, biological evaluation, and modeling of a non-carbohydrate antagonist of the myelin-associated glycoprotein."

I performed all the molecular simulations of this project with the exception of the MAG homology model, which was provided by Dr. Morena Spreafico.

Publication 3: S.Mesch et al. "Low Molecular Weight Antagonists of the Myelin-Associated Glycoprotein: Synthesis, Docking, and Biological Evaluation."

I performed all the molecular simulations of this project with the exception of the MAG homology model, which was provided by Dr. Morena Spreafico.

Publication 4: G. Rossato et al. "AcquaAlta: A Directional Approach to the Solvation of Ligand-Protein Complexes."

This project both in term of planning, simulations, calculations, and analysis was completely carried out by me. The *ab initio* calculations were performed by Dr. Martin Smiesko.

Table of Contents

AIM OF THE THESIS	13
INTRODUCTION	15
1 WATER IN MOLECULAR SIMULATIONS AND DRUG DESIGN	15
1.1 <i>Water in drug design</i>	15
1.1.1 Water models.....	17
1.1.2 Factors influencing water treatment and evaluation in X-ray crystal structures.....	19
1.1.2a Water stability related to hydrogen-bond quality and counts.....	20
1.1.2b Isotropic thermal displacement factors of water molecules (B _{iso}).....	21
1.1.2c Water accessibility and hydrogen-bond networks.....	22
1.2 <i>Approaches to hydration in drug design</i>	24
1.3 <i>AcquaAlta: Solvation of ligand–protein interfaces</i>	25
2 METOX: METABOLISM AND TOXICOLOGY PREDICTION	27
2.1 <i>Cytochromes P450 and Metabolism</i>	27
2.1.1 Different Isoforms, concentration and contribution.....	28
2.1.2 Oxidation states and factors affecting CYP activity	29
2.1.3 Pharmacophore models.....	30
2.1.4 Cytochromes flexibility / promiscuity	32
2.2 <i>Approaches in cytochrome modeling: 3D-QSAR</i>	33
2.3 <i>Mixed-model QSAR at cytochrome P450 2D6 and 2C9</i>	35
3 COMPUTER-AIDED DRUG DESIGN AT THE MYELIN ASSOCIATED GLYCOPROTEIN (MAG)	37
3.1 <i>Axonal regeneration in adult mammalian CNS</i>	37
3.2 <i>MAG: Myelin-Associated Glycoprotein</i>	39
3.3 <i>Design of MAG antagonists</i>	40
METHODS	43
1 WATER IN MOLECULAR SIMULATIONS AND DRUG DESIGN	43
1.1 <i>Cambridge Structural Database search</i>	43
1.2 <i>Ab initio calculation</i>	43
1.3 <i>Organization of the algorithm</i>	44
1.4 <i>Validation of AcquaAlta</i>	48
1.5 <i>Application of AcquaAlta: Docking with Alignator</i>	50
2 METOX: METABOLISM AND TOXICOLOGY PREDICTION	51
2.1 <i>Protein structures</i>	51
2.2 <i>Ligand structures and affinity data</i>	53
2.3 <i>Prealignment and automated, flexible docking</i>	53
2.4 <i>Multidimensional QSAR</i>	56
2.5 <i>Virtual screening applied to toxicology prediction (Virtual Tox Lab)</i>	57
3 COMPUTER-AIDED DRUG DESIGN AT THE MYELIN ASSOCIATED GLYCOPROTEIN (MAG)	59
3.1 <i>Docking to the MAG homology model</i>	59

RESULTS	63
1 WATER IN MOLECULAR SIMULATIONS AND DRUG DESIGN	63
1.1 <i>Hydrogen-bond directionality (in water networks)</i>	63
1.2 <i>Functional-group hydration propensity</i>	70
1.3 <i>Validation of AquaAlta</i>	72
1.4 <i>Algorithm application: Docking with Alignator</i>	77
1.5 <i>Comparison with the solvation module of Yeti</i>	80
2 METOX: METABOLISM AND TOXICOLOGY PREDICTION	82
2.1 <i>Correlation of experimental and calculated atomic motion</i>	82
2.2 <i>Prealignment and automated, flexible docking</i>	85
2.3 <i>Multidimensional QSAR (mQSAR)</i>	89
2.4 <i>Consistency with the sites of metabolism</i>	91
2.5 <i>Y-scrambling and binding of external datasets to CYP2D6 and CYP2C9</i>	93
2.6 <i>VirtualToxLab dataset screening</i>	98
3 COMPUTER-AIDED DRUG DESIGN AT THE MYELIN ASSOCIATED GLYCOPROTEIN (MAG)	99
3.1 <i>Experimental data and molecular docking</i>	99
3.1.1 Series 1.....	99
3.1.2 Series 2.....	101
3.1.3 Series 3.....	103
3.1.4 Series 4.....	108
3.2 <i>Molecular dynamics simulations</i>	109
3.2.1 Series 1.....	109
3.2.2 Series 2.....	112
3.2.3 Series 3.....	113
3.2.4 Series 4.....	118
3.3 <i>Estimation of relative-free energy differences using FEP</i>	122
3.3.1 Series 3.....	122
CONCLUSIONS	123
1 WATER IN MOLECULAR SIMULATIONS AND DRUG DESIGN	123
2 METOX: METABOLISM AND TOXICOLOGY PREDICTION	125
3 COMPUTER-AIDED DRUG DESIGN AT THE MYELIN ASSOCIATED GLYCOPROTEIN (MAG)	127
REFERENCES	129
APPENDIX 1: SUPPORTING INFORMATION OF “WATER IN MOLECULAR SIMULATIONS AND DRUG DESIGN”	145
APPENDIX 2: CHEMMEDCHEM 2010	160
APPENDIX 2A: CHEMMEDCHEM 2010 – SUPPORTING INFORMATION	175
APPENDIX 3: BIO. & MED. CHEM. 2010	195
APPENDIX 4: J. MED. CHEM. 2010	209
APPENDIX 5: J. CHEM. INF. MOD. 2011	209

Abbreviations

CNS	Central Nervous System
CPU	Central Processing Unit
CSD	Cambridge Structural Database
CYP	Cytochrome P450
e.g.	Exempli gratia (for example)
FEP	Free Energy Perturbation
FMI	Flavin-Containing Monooxygenase
HBA	Hydrogen-bond acceptors
HBD	Hydrogen-bond donors
HEV	Hydrogen extension vector
HMG-CoA	3-hydroxy-3-methyl-glutaryl-CoA reductase
i.e.	Id est (that is)
IC ₅₀	Inhibitory Concentration causing 50% inhibition
IG	ImmunoGlobulin
K _i	Inhibition Constant
ITC	Isothermal Calorimetry
LIE	Linear Interaction Energy
LOO	Leave-One-Out cross validation
LPV	Lone pair vector
MAG	Myelin Associated Glycoprotein
MAO	Monoamine Oxidase
MC	Monte Carlo
MD	Molecular Dynamics
MLR	Multiple Linear Regression

MM-PBSA	Molecular Mechanics-Poisson Boltzmann Surface Area
NAT	N-AcetylTransferase
Neu5Ac	5-N-acetyl neuraminic acid
NCAA	National Collegiate Athletic Association
NgR	Nogo Receptor
NSAIDs	Non-Steroidal Anti-Inflammatory Drugs
Omgp	Oligodendrocyte myelin glycoprotein
OppA	Oligopeptide binding Protein A
p75 ^{NTR}	P75 neurothropin receptor
ρ^2	Predictive correlation coefficient
PCA	Principal Component Analysis
PCR	Principal Component Regression
PDB	Protein Data Bank
PNS	Peripheral Nervous System
q^2	Cross-validated correlation coefficient
QSAR	Quantitative Structure Activity Relationship
r^2	Correlation coefficient
<i>rmsd</i>	root mean square deviation
SIGLECS	Sialic acid binding ImmoGlobuline –like LECTins
SPR	Surface Plasma Resonance
TI	Thermodynamic Integration
UGT	Uridine Dinucleotide Phosphate
UPT	Glucuronosyl Transferase
VTL	VirtualToxLab

Aim of the thesis

This thesis focuses on the use of computational techniques (*i.e.* *in silico*) applied to different pharmacological targets and phases of the drug discovery process. The use of predictive computational tools represents an important aid in drug discovery by speeding up research, assisting in the selection of potential drug candidates, in the lead optimization phase, and in the optimization of both pharmacodynamics and pharmacokinetic profile. *In silico* techniques are often fast and cheap in terms of resources needed, and if validated, they can be considered as potential replacements for *in vitro* and *in vivo* assays.¹ An increasing number of validated examples showed how the discovery of new drugs is often the results of fruitful combinations between experimental and computational techniques.

This thesis is divided in three distinct sections, organized in their separate introductions, methods, results, and conclusions. The three sections are entitled: (1) Water in molecular simulations and drug design; (2) MeTOX: Metabolism and TOXicology prediction; (3) Computer-aided drug design on the Myelin Associated Glycoprotein (MAG).

– *Water in molecular simulations and drug design*: Despite the ubiquitous presence of water molecules in X-ray structures and their known importance for a reliable calculation of binding affinities, desolvation cost along with entropy changes upon ligand binding in computational drug design, their correct accounting is still far from being solved. Here, we present an approach for the solvation of ligand–protein complexes along with its validation and external application.

– *MeTOX: Metabolism and TOXicology prediction*: Cytochrome P450 are enzymes involved in the metabolism of the majority of the marketed compounds. A multidimensional QSAR study on CYP450 2D6 and 2C9 was performed. The models were subsequently used to quantify binding affinities for external dataset of compounds (both binders and not-binders). The validated models were finally implemented in the VirtualToxLab (www.virtualtoxlab.org)

– *Computer-aided drug design on the Myelin Associated Glycoprotein (MAG)*: Docking, molecular dynamics, and free energy perturbation simulations were performed on the Myelin Associated Glycoprotein to unveil the structural and mechanistic aspects of the inhibition of this receptor involved in axon regeneration.

Introduction

1 Water in molecular simulations and drug design

1.1 Water in drug design

In nature, and in particular at a molecular level, vacuum is generally avoided. Endogenous ligands and xenobiotics binding to biomolecules are typically solvated and they utilize solvent molecules during their molecular recognition events. Therefore, any ligand binding to any biomolecule displaces some water molecules from the binding site, while undisplaced water molecules either bridge intra- (ligand–ligand and protein–protein) and intermolecular (ligand–protein) interactions or fill empty polar and apolar cavities.^{2, 3} Waters bridging polar interactions at the binding site can therefore improve both specificity of interactions during binding together with the relative binding affinity.⁴

Water molecules also modify the flexibility of biomolecules and can enhance complementarity between protein and ligand.⁵⁻¹¹ Water molecules are additionally involved in catalytic enzymatic reactions and in the characterization of the solvation/desolvation profile of each ligand and bio-macromolecule.

An analysis of 392 high-resolution protein complexes, retrieved from the Protein Data Bank (PDB),¹² showed that over 85% of the protein-ligand complexes display one or more water molecules bridging ligand and protein. The average number of ligand-bound water molecules was found to be 4.6 and 76% of these water molecules were identified engaging in polar interactions with both ligand and protein.¹³

Water is a molecule, acting both as hydrogen-bond acceptor and hydrogen-bond donor, and is smaller than any protein polar side chain, and ubiquitous in most molecular environments.¹⁴ It is highly polar and polarizable, and it is able to interact with other solvent molecules establishing different kinds of water networks (e.g. hydrogen-bond networks, proton relay systems).

One of the objectives of computer-aided drug design is a quantitative estimation of free energies of binding.¹⁵ The free energy of binding is a thermodynamic

entity characterized by the sum of an enthalpic and entropic contribution. The binding is driven by free energy changes associated to the interactions involved in the binding.¹⁶ When estimating free energies of binding we basically subtract two large numbers: (i) the ligand–protein interaction energy and (ii) the hydration energy.¹⁷ Both of the estimations are associated to errors and furthermore the computation of reliable hydration energies are challenging because they should consider the differences in entropy of the solvent after ligand binding. The hydration energy should assemble solvation, desolvation and hydrophobic solvation terms upon binding.¹⁸

When we consider the hydration energy, each water molecule presents a different entropy/enthalpy balance. A highly mobile water will cause a high entropy when immobilized in a certain position, which could be balanced from the enthalpy gain contributed by water-mediated hydrogen bonds.¹⁹

The release of water molecules (stably bound to a protein) to the bulk is entropically favorable, and the entropy gain to transfer a stable water molecule from the protein to the solvent is estimated to be of 2 kcal/mol at room temperature.²⁰

Further studies, using a double-decoupling method,²¹ estimated the free energy necessary to remove a water molecule from the binding site of a ligand–protein complex to be of -1.9 ± 0.5 kcal/mol for trypsin/benzylamine complex and -3.1 ± 0.6 for the HIV/KNI-272 complex.

Substantial efforts aiming at categorizing water molecules as stable (often defined as structural) or displaceable were done. Structural water molecules establishing three or more hydrogen bonds,²² are influenced by local site surface shape preferring buried cavities,²³ and are conserved at ligand-binding sites of homologous proteins.²⁴ In addition, buried water molecules present a long residence time (10^{-8} – 10^{-2} s).¹⁹ From the structural point of view, binding sites can be analyzed by solvent mapping²⁵ or using modular neural networks.²⁶ Based, on these information, the role of the identified water molecules can be investigated aiming at the understanding if a more favorable interaction can be gained with a ligand substitution which replace a water molecule, or if a more favorable binding can be reached keeping a water molecule in the original place.

Different studies²⁷⁻³¹ suggest that a replacement of binding site waters by additional ligand moieties lead to either more potent or equipotent inhibitors. In other cases,³²⁻³⁴ the presence of water molecules and the correspondent binding stabilization was difficult to mimic, leading to a decreased binding affinity.

1.1.1 Water models

For the aspects listed in chapter 1.1, water is one of the major research subjects both before and after the advent of computer simulations in the late 1960. Some of the milestones which characterized the studies on waters are the following.³⁵

- 1932 - Spectroscopic proofs of the V-shaped water molecule
- 1933 - Realistic interaction potential for water
- 1953 - Monte Carlo sampling scheme
- 1957 - Molecular dynamics (MD) simulation
- 1969 - Computer simulations of water
- 1976 - Pair potential from *ab initio* calculations for water
- 1981 - "Accurate" and simple pair potential for liquid water
- 1993 - *Ab initio* calculations for liquid water

In order to correctly simulate water at a molecular level, the choice of which water model to use is fundamental. Water models are evaluated based on their ability to reproduce the properties of real water, in different conditions of temperature and pressure. Seven different properties are object of evaluation: the density of the liquid, the heat of vaporization, the self-diffusion coefficient, the atom–atom pair distribution functions, the temperature of maximum density, the critical parameters and the dielectric constant.³⁵ The presence of up to 67 water anomalies reflects the high complexity of its correct simulation.³⁶

Three main types of water models are used, namely: (i) rigid models, (ii) flexible models, (iii) polarizable models. Flexible models compared to rigid models, where water is considered as a rigid body, include bond stretching and H-O-H angle bending, and it can reproduce the vibration spectra. Polarizable water models include explicit polarization terms, which should increase the reproducibility of interactions among water molecules and of different water phases.

Focusing on rigid, non-polarizable water models, the most widely used models are: SPC,³⁷ SPC/E,³⁸ TIP3P,³⁹ TIP4P³⁹ and TIP5P.⁴⁰ These models differ in bond geometries, charge distribution (i.e. location of the negative charge(s)), and the properties used to force the model to reproduce the experimental properties (Lennard Jones potential σ and ϵ , the value of charge on the hydrogen atoms). A comparison of the Lennard-Jones potential for the listed water models is depicted in Figure 1.

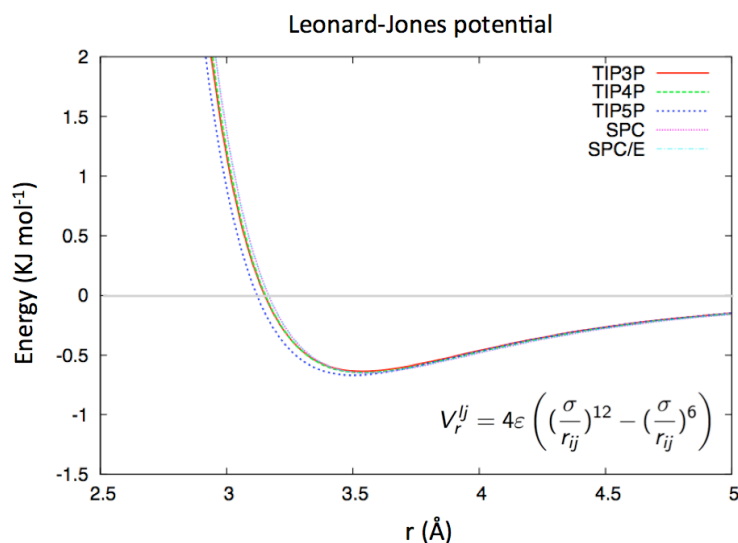


Figure 1: Lennard-Jones potential expressed in kJ/mol for five different water models, namely: TIP3P, TIP4P, TIP5P, SPC, SPC/E. The figure is adapted from Caleman et al.⁴¹

Hereafter are summarized the different properties which characterize some of the rigid water models:

- SPC: The bond length is 1 Å and the H-O-H angle is 109.47°. The parameters of the model are aiming to reproduce the enthalpy of vaporization and the liquid density at room temperature.
- SPC/E: Identical to the SPC model, but it reproduces the vaporization enthalpy of experimental water when a polarization energy correction is included.
- TIP3P: The bond length is 0.9572 Å and the H-O-H angle is 104.52°. The negative charge is on the oxygen atom and the positive charges on the hydrogen atoms. Charges: –0.8 for the oxygen atom and 0.4 for each hydrogen atom.
- TIP4P: The negative charge is localized on the H-O-H bisector at a distance of 0.15 Å. Charges: –1.04 at 0.15 Å of the H-O-H bisector, and 0.52 on each hydrogen atom.
- TIP5P: The negative charge is localized at the two positions of the lone pair electrons. Compared to other models, TIP5P introduces the match of the target property of the maximum density of liquid water.

A comparison of the experimental and reproduced physical properties is listed in Table 1.

Table 1: Physical properties for experimental and different water models. The physical properties listed are the dipole moment in the gas phase (μ_{Gas}), in the liquid phase (μ_{Liq}), the conformational energy (E_{conf}), the dielectric constant and the force-field in which the different water models were produced.

Model	$\mu_{\text{Gas}}(\text{D})$	$\mu_{\text{Liq}}(\text{D})$	$E_{\text{conf}}[\text{kcal/mol}]$	ϵ	Force field
Experimental	1.855	2.95 (298K)	-9.91	78.4	—
SPC	2.27	2.27	-9.79	65	GROMOS
SPC/E	2.35	2.35	-9.91	71	GROMOS
TIP3P	2.35	2.35	-9.84	82	AMBER
TIP4P	2.18	2.18	-9.98	53	OPLS
TIP5P	2.29	2.29	-9.86	81.5	OPLS

1.1.2 Factors influencing water treatment and evaluation in X-ray crystal structures

When quantifying the energy of water molecules at the binding interfaces, we often encounter with the questions: where are located in the space these water molecules, and how long is their interaction time. Information about position with respect to the hydrogen-bond partners involved in the interaction, and water residence time will translate in the strength of interaction with the protein as well as to the contribution in ligand–protein binding, and protein stability. The most common methods to obtain structural information on biocomplexes are four: X-ray crystallography, NMR spectroscopy, electron microscopy and computer methods. These detection models generally identify stable and therefore ordered water molecules.

In X-ray structures, water molecules can be distinguished at high resolution. Typically, structures with resolution better than 2.0 Å are considered as highly resolved and they allow a resolution of stable water molecules. In X-ray structures, each water molecule is identified by an electron-density, which is determined along the period of data collection. The final water position is assigned through the identified electron-density.⁴² Waters are identified only in the last refinement steps, and their presence (due also to the high number of

solvent molecules) can substantially help in improving the fit to the data and the corresponding statistical parameters (e.g. R factor), which mirror the quality of a crystal structure. For these reasons, each water molecule should be considered with caution,⁴³ and with the respect to the other structural evidence collectable from the structure (e.g. the localization of hydrogen-bond acceptors and hydrogen-bond donors around the water molecule).

Factors influencing and reflecting the presence, localization, interactions pattern, and role of each water molecule among others are:

- Water stability related to hydrogen bond quality and counts
- Thermal factor of water molecules in the X-ray structures
- Water accessibility and hydrogen-bond networks

Each of these factors will be treated separately in the next chapters. Structurally important waters are also found conserved in homologous proteins highlighting their importance in the understanding of the molecular mechanisms of evolutionally related proteins.²⁴

1.1.2a Water stability related to hydrogen-bond quality and counts

The identification of water molecules in structures collected from the PDB, does not directly correlate with the stability and structural role of these solvent molecules. Water displays an almost regular tetrahedral coordination and its potential of acting both as hydrogen-bond acceptor and hydrogen-bond donor allows it to form up to four strong hydrogen bonds (i.e. two with the water hydrogens and two with the free lone pairs of the water oxygen.).

A hydrogen bond is a strong electrostatic interaction between an electronegative and a hydrogen atom bonded to an electronegative atom. The strength of a hydrogen bond is defined both by its interaction energy and by the geometric criteria defining it. Geometric criteria have been obtained from analyzing small-molecule crystal structures where the positions of hydrogen atoms are observable.⁴⁴⁻⁴⁹ Similarly (cf. Results – Section 2) we searched the Cambridge Structural Database (CSD) to identify the geometrical criteria defining interactions of water molecules with ligand and protein. On these bases, a hydrogen bond (Don–H \cdots Acc) must fulfill three geometrical criteria:

- *Distance*: The distance between the donor and the acceptor must be below the sum of their individual van der Waals radii (e.g. O–O: 3.04 Å, O–N: 3.07 Å, O–S: 3.32 Å).
- *Linearity*: Water is seeking to establish hydrogen bonds with a linear Don–H···Acc arrangement.
- *Directionality*: Hydrogen bonds are usually directed along the lone pair(s) of the acceptor atom.⁴⁴ A deviation (maximally up to 45°) from this arrangement is observable and acceptable.

The number of simultaneous hydrogen bonds established also influences stability of water in X-ray structures. Nearly the 80% of the water molecules bridging ligand–protein interactions can establish three or more hydrogen bonds with ligand, protein and other solvent molecules.²²

1.1.2b Isotropic thermal displacement factors of water molecules (B_{iso})

The isotropic thermal displacement factor (also called Debye-Waller factor or isotropic B factor) is a measure of how much an atom oscillates around its average position (i.e. the position in the model). The isotropic B factor (Table 2) reflects the mean displacement (\bar{u}) of an atom about its mean position and it is a direct measurement of the effective mobility of a certain atom.

Table 2: The thermal factor (B factor) of an atom (j) expressed as function of the mean square displacement (\bar{u}^2). Examples of B factor values along with their correspondent displacements in Å are listed in the right part of the table.

$$B_j = 8 \pi^2 \langle \bar{u}_j \rangle^2$$

B	\bar{u}
79.0	1.0 Å
20.0	0.5 Å
5.0	0.25 Å

Waters establishing more than two hydrogen bonds show a decreased B-factor, therefore their mobility, compared to solvent molecules where the number of hydrogen bonds is lower, is also lowered.²² Despite a higher number of hydrogen bonds (i.e. a more stable water molecule) should translate in lower B-factor values, this is not always true as these values are often subjected to normalization. To correctly interpret these data, a rigorous analysis should also

consider the thermal factor values of the surrounding atoms. Generally, water molecules with thermal factor values in the same range, or even lower than protein atoms are considered as highly ordered waters, and are defined as structural waters.

1.1.2c Water accessibility and hydrogen-bond networks

The solvent accessibility of the binding site (Figure 2), is an important aspect to consider prior to the use of methods/software to solvate or to evaluate solvation effects.

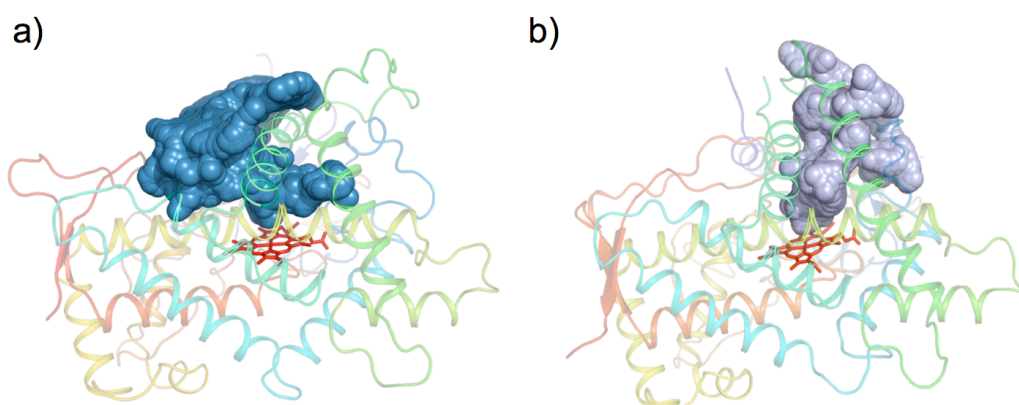


Figure 2: Solvent accessibility of the active sites of CYP2D6 (a) and CYP2C9 (b) produced with HOLLOW⁵⁰ and rendered with PyMol.⁵¹

Typically in proteins, the binding sites are either permanently or temporarily solvent accessible. The solvent accessibility of the binding site allows for exchange of buried water molecules with bulk waters. For the bovine trypsin inhibitor, this has been estimated to be on the scale time of 15 ns to 1 μ s.⁵² Nonetheless, it is not infrequent to identify cavities, which display no contact with the bulk solvent. Waters can be inserted in small apolar cavities, which minimum size was calculated to be of about 500 \AA^3 , allowing clusters of three or four water molecules.³

Indeed, water networks represent stable ensembles, both when located in small cavities and in bulk water. They are characterized by recurrent geometries as well as by high energies. When designing new ligands the cost for disorganizing or breaking such high-energy entities should be accounted. An example of a

water network as simulated in a MD simulation using the TIP3P water model is depicted in Figure 3.

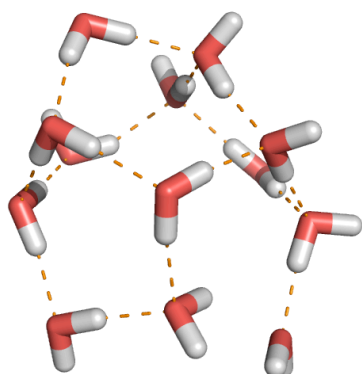


Figure 3: Example of water network using the TIP3P water model. The hydrogen bond network is shown with dashed lines.

In liquid water, water molecules undergo a continuous protonation and deprotonation process. As a consequence hydrogen bonds are continuously established and broken in a cooperative way. Several theories have been proposed for the structure of these water clusters, from random network models⁵³ to the icosahedral water cluster model (Figure 4).⁵⁴ These models are generally evaluated based on their ability to simulate the higher number of water physical properties and anomalies.⁵⁵

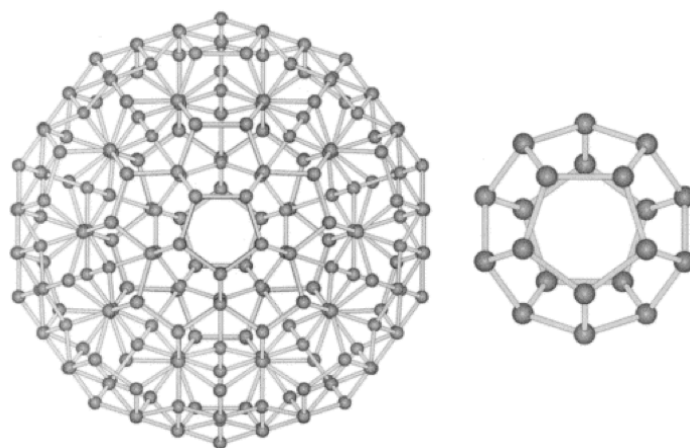


Figure 4: Figure adapted from Chaplin et al.⁵⁴ depicting 280 water molecules organized in an icosahedral water cluster. On the right only the cluster of water oxygen atoms is shown.

1.2 Approaches to hydration in drug design

Hydration and its effects on ligand binding has been addressed in several studies using different computational techniques.

The free energy of binding can be calculated using free energy perturbation (FEP) or thermodynamic integration (TI). These methods are able to predict free energies of binding within 1 kcal/mol, but they are found to be reliable only when estimating relative energies (i.e. comparing chemically similar molecules).¹⁸

These methods are computationally demanding, therefore novel approximation methods aim to obtain a compromise between prediction accuracy and time of computation. Approximate methods such as molecular mechanics-Poisson Boltzmann surface area (MM-PBSA) or linear interaction energy (LIE), extrapolate free energies of binding from MD simulations for the MM-PBSA method and from MD or Monte Carlo (MC) simulations for the LIE method.

Many predictive tools were developed to characterize and distinguish the role of water in biomolecules.^{56,57} Potential hydration sites were evaluated by solvent mapping or using (modular) neural networks.^{25 26}

Water molecules are treated either implicitly (by continuum solvation methods), or explicitly.⁵⁸⁻⁶¹ When treated explicitly, they can be considered in molecular docking, and it has been shown how ordered water molecules can actively influence ligand recognition.⁶² Some docking programs allow switching “on” and “off” the presence of solvent molecules originating from X-ray structures (adopted approaches in different docking programs are listed in Table 3).^{58, 63}

Nowadays many efforts are directed to re-parameterize scoring functions used in molecular docking to account for the interactions of these water molecules and/or for the desolvation cost of each ligand-binding event.

More recently (2008), the contribution in terms of enthalpy/entropy balance of explicit water molecules is estimated using free energy methods based on MD simulations.⁶⁴

Despite all these approaches, the real value of considering water molecules in molecular simulations rises controversial opinions. Some studies emphasize the need of water molecules to improve accuracy in molecular docking,^{55, 65} while others suggest that the presence of water improved the docking results only marginally.⁶⁶

Table 3: Currently adopted approaches in different docking programs.

Software name	Approach
<u>Cheetah</u> www.biograf.ch	Dynamic displacement of waters from a presolvated uncomplexed binding site
<u>FlexX</u> – Particle Concept www.biosolveit.de/flexx/	“Flexible water generation”. Waters are suggested for each ligand generation
<u>GLIDE</u> www.schrodinger.com	Position to retain selected water molecules during the docking runs
<u>GOLD</u> www.ccdc.cam.ac.uk/	Automatic keeping and retaining of water molecules during the docking run
<u>MOE</u> – Docking www.chemcomp.com	Position to retain selected water molecules during the docking runs
<u>Molegro Virtual Docker</u> www.molegro.com	Automatic keeping and retaining of water molecules during the docking run generated by the software based on ligand atom positions

1.3 AcquaAlta: Solvation of ligand–protein interfaces

In this thesis, a novel approach (AcquaAlta) has been developed. It generates explicit water molecules at ligand–protein interfaces. The underlying algorithm relies on geometric preferences for water location and orientation as extracted from structural information collected from the Cambridge Structural Database — CSD (currently including the structures of more than 500,000 organic molecules). Specifically, water molecules interacting with generic functional groups (e.g. the carbonyl query is comprehensive of both carboxylic acid, esters and amides groups) of organic small molecules were searched. To establish a hydration-propensity ranking, interaction energies were obtained using *ab initio* calculations on hydrated functional groups. In AcquaAlta, water molecules bridging interactions between ligand and protein partners are generated considering these calculated hydration propensities of the involved functional groups and aromatic residues.

The presence of a tool, such as AcquaAlta, aiming to rationally find bridging waters in protein complexes, can represent an aid, in the correct scoring of binding poses resulting from diverse docking protocols. The presence of waters in the binding site can allow to identify bridged interactions, which contrarily they would be neglected. Moreover, along with a possibly more reliable estimation of the binding affinity the presence of water molecules placed in the correct position, can minimize ligand movements during the refinement phase (e.g. during the minimization of the complex).

For the validation of this concept, AcquaAlta was used to reproduce bridging water molecules found in 20 mainly high-resolution protein crystal structures. The structures belong to different target families and are listed together with their correspondent properties (i.e. PDB code, resolution, R-factor and number of binding site waters) in the methods section 1.4.

AcquaAlta was then applied to the docking of oligopeptides binding to the periplasmic oligopeptide binding protein (OppA). The oligopeptides were docked using the pharmacophore-based docking tool Alignator. The solvation algorithm was applied to both crystal poses and docked poses and the water produced were compared to the experimental water molecules at the ligand–protein interface of the respective structures. The accuracy of the docking was evaluated calculating the *rmsd* between crystal and docked poses. The accuracy of the solvation algorithm was evaluated checking the match between experimental waters and calculated waters based both on crystal and docked poses.

We selected OppA because a large number of highly resolved ligand-protein complexes is available. The structural similarity of the complexed oligopeptides makes it suitable for the application of the selected docking protocol with Alignator. Interestingly, different hydration patterns among *apo* and ligand-bound *holo* OppA have been reported.

2 MeTOX: Metabolism and TOXicology prediction

2.1 Cytochromes P450 and Metabolism

In drug discovery and development, drug metabolism and safety are gaining increasing attention. Late-stage failures are frequently related to toxicology problems. For this reason, the paradigm “fail early, fail cheap” is showing where the efforts of pharmaceutical research are directed to.

Manifestation of toxic effects is connected to cascades of events, often influenced by the human metabolism. The metabolism induced by endogenous enzymes such as the cytochrome P450 enzymes (CYPs) is of major importance for the degradation and excretion of drugs.⁶⁷ An important functional role of CYPs is the synthesis of signaling factors for controlling development and homeostasis.⁶⁷ For example, CYPs convert polyunsaturated fatty acids to active molecules and convert steroid hormones and fat-soluble vitamins.⁶⁷

CYPs are hemoproteins involved in the phase-1 metabolism. The phase 1 metabolism includes hydroxylation, epoxidation, N or S oxidation and dealkylation processes. During this phase the lipophilicity of xenobiotics is decreased, through oxidative reactions, in order to increase the clearance of the metabolites.

The family of cytochromes P450s, takes its name based on the particular absorption peak at 450 nm with carbon monoxide-bound pigment,⁶⁸ where P refer to the pigment and 450 to the wavelength of maximum absorbance. CYPs are found in mammals, plants, fungi and bacteria, and in mammals, their distribution is prevalent in liver tissues followed by lungs and skin localization.

CYPs contribute to the metabolism of the largest amount of marketed drugs together with the enzymatic reactions catalyzed by uridine dinucleotide phosphate (UGT), glucuronosyl transferase (UDP), flavin-containing monooxygenase (FMO), *N*-acetyltransferase (NAT), and monoamine oxidase (MAO).⁶⁹

2.1.1 Different Isoforms, concentration and contribution

Cytochromes comprehend almost 500 genes,⁷⁰ which are classified in families and subfamilies. In the common cytochrome nomenclature, families are referred by Arabic numbers, subfamilies by letters, and the isoenzyme type by the last Arabic number. The subfamilies involved in the drug metabolism are the following: 1A, 1B, 2A, 2B, 2C, 2D, 2E, and 3A and the concentration of each of the isoform in the hepatic system is different as shown in Table 4.

Table 4: Adapted list⁷¹⁻⁷³ of the most relevant CYP isoforms and their liver concentration in human bodies.

CYP isoforms	Human liver concentration
3A4 / 3A5	~ 30 %
2C8 / 2C9 / 2C18	~ 20 %
1A2	~ 15 %
2E1	~ 10 %
2D6 / 2C19 / 2A6	< 5 %
1B1 / 2B6	~ 1 %

Nonetheless, the concentration of the different isoforms in the liver does not directly correlate with the relative contribution of these isoforms in the interaction with drugs (Figure 5). CYP 1A2, 2C9, 2C18, 2C19, 2D6 and 3A4 contribute to the catalytic reactions of approximately 85% of the marketed compounds.

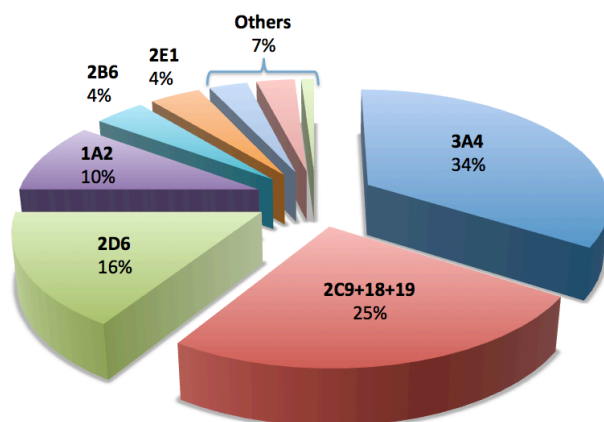


Figure 5: Chart (adapted from Rendic et al.⁷³) showing the contribution of the different human CYP isoforms in the interaction with marketed drugs (~6200, Source: Drugs@FDA – <http://www.accessdata.fda.gov/scripts/cder/drugsatfda/index.cfm>.)

2.1.2 Oxidation states and factors affecting CYP activity

The catalytic mechanism of P450 (Figure 6) can be represented by a series of chemical steps requiring redox partners and the use of reducing equivalents in the form of nicotinamide-adenine dinucleotide phosphate (NADPH) or reduced nicotinamide-adenine dinucleotide (NADH).⁶⁷ The reaction cycle contains steps involving the activation of an oxygen molecule, the oxidation of a substrate and the release of the (newly formed) product.

After substrate binding (a), one electron from NADPH or NADH is transferred to the heme-bound iron Fe^{3+} (b); molecular oxygen binds to the reduced heme iron and there is the formation of an oxygenated heme $\text{Fe}^{2+}\text{-O}_2$ or $\text{Fe}^{3+}\text{-O}_2$ state (c). A second electron reduces this complex to a ferric peroxo state $\text{Fe}^{3+}\text{-O}_2^-$ (d), which is easily protonated to form the hydroperoxo $\text{Fe}^{3+}\text{-O}_2\text{H}$ state (e). A second protonation of the latter complex at the distal oxygen atom allows the heterolytic scission of the O–O bond from the unstable state $\text{Fe}^{3+}\text{-O}_2\text{H}_2$ and the release of a water molecule (stage f). The porphyrin-oxo-complex oxidates the substrate, allowing its release (stage g) and the heme-bound iron turns back to the original Fe^{3+} state.⁶⁷

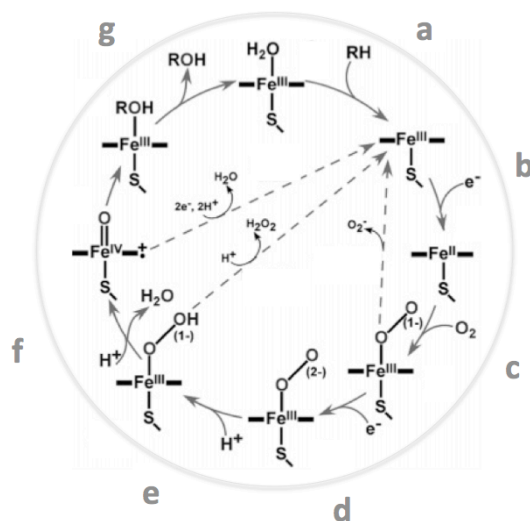


Figure 6: The cytochrome P450 reaction cycle Adapted from Denisov et al.⁶⁷

Three main factors can influence the activity of CYPs, which are: induction, inhibition and pharmacogenetics.⁷⁴

– *Induction* refers to molecules, which enhance the expression of an enzyme. Some studies⁷⁵ suggest that barbiturates (e.g. phenobarbitone) can induce CYP2C9 activity. This is deduced from the diminished steady-state plasma concentration of S-warfarin and the correspondent anticoagulant effect⁷⁶ when co-administred with phenobarbitone. In turn this suggests an induction of the S-warfarin metabolism.

– *Inhibition* refers to the inhibition of the expression of the enzyme by the interaction with a molecule or by competitive, non-competitive, or through a partially competitive inhibition. For example, Fluconazole, together with other azole antifungals inhibits CYP2C9 both *in vitro* and *in vivo*.⁷⁷ Similarly, the HMG-CoA reductase inhibitor Fluvastatin is a potent inhibitor of the hepatic CYP2C9 activity *in vitro*.⁷⁸

– *Pharmacogenetics* refers to the genetic variation, which affects the response to a particular drug. The polymorphism of CYPs is the major factor, which influences the different clearance/response to drugs by different individuals. A diminished activity of the CYP2D6 is found in 5–14% of the Caucasians, 0–5% of the Africans and 0–1% of the Asians.⁷⁹ A high polymorphism is found also for CYP2C9, which modifies the response to drugs like coumarin anticoagulants, sulfonylurea drugs, angiotensin II inhibitors, phenytoin and non-steroidal anti-inflammatory drugs (NSAIDs).⁷⁹

2.1.3 Pharmacophore models

Despite the high variety of structurally diverse compounds binding to the various isoforms involved in intestinal and hepatic drug metabolism (1A2, 3A4, 2B6, 2D6, 2C9, 2C18, 2C19, 2E1), a pharmacophore preference from the different isoforms is observable. Pharmacophore models can potentially be used to model binding and metabolism of compounds on isoforms where structural information are lacking. This in turn gives an indirect information on the active site and can be used to establish structure-activity relationships (SAR) on structurally similar compounds binding to the same isoform.

Similar structural features from different ligands are compiled in ligand-based pharmacophore models, aiming to identify chemical similarities (i.e. pharmacophoric groups), from which it is possible to derive a pharmacophore

hypothesis. Those hypotheses can be used for a fast ligand-based screening, as well as for a search of compounds displaying similar pharmacophoric features.

In-depth studies have been carried out to generalize pharmacophore hypothesis for the different cytochrome isoforms.⁸⁰ In Table 5, there are listed common features of ligands binding (as substrates or inhibitors) to CYP 1A2,⁸¹ 3A4,⁸² 2B6,⁸³ 2D6,⁸⁴ 2C9,^{80, 85} 2C18, 2C19,⁸⁶ 2E1.⁸⁷

Table 5: List of the pharmacophore hypotheses for the major CYPs isoforms involved in intestinal and hepatic metabolism. Each hypothesis is as coming from the references above.

CYP isoform	Pharmacophore hypothesis
1A2	Substantial hydrophobic interactions between ligand and protein. An area presenting negative charge is close to the nitrogen atom of the quinolone
3A4	Inhibitor model: Three hydrophobic regions at 5.2–8.8 Å from an HBA, three hydrophobic regions at 4.2–7.1 Å from an HBA and an additional 5.2 Å from another HBA or one hydrophobic region at 8.1–16.3 Å from 3 HBAs
2B6	One hydrophobic and one HBA functionality. Two different studies underline the presence either of two or three hydrophobic regions
2D6	Substrates model: a positively charged nitrogen, an HBA and a hydrophobic region. The site of oxydation is 5–7 Å away from the site of oxydation. Inhibitor model: an HBA and an HBD and two or three hydrophobic regions.
2C9	Inhibitors have one hydrophobic moiety and one HBA. HBA are at 3.4-5.7 Å distance from a HBD or an additional HBA. The hydrophobic feature is located at 3.0-5.8 Å from the HBA
2C18 2C19	An electropositive region and two hydrophobic sites.
2E1	Presence of many hydrophobic regions due to the long hydrophobic access channel.

2.1.4 Cytochromes flexibility / promiscuity

A peculiarity of CYPs is their highly promiscuous character (Figure 7). They are able to bind a large variety of different substrates, therefore CYPs feature a high flexibility of their active site.⁸⁸

For both CYP3A4 and CYP2C9, the structures of apo and holo-enzyme are available in the PDB. For CYP3A4 the protein conformation of the ligand-bound structures is substantially different from the structures of the apo-enzymes. In the case of CYP3A4 the substrate progesterone⁸⁹ was located in a part of the active site far from the iron atom of the heme, and therefore far from the site of oxidation. Similarly warfarin bound to CYP2C9⁹⁰ is located at a distance of 10 Å from the site of oxidation.

The complex of ketoconazole with CYP3A4, when superimposed to the ligand-free structures yields a remarkable high *rmsd* of 1.6 Å for the C α atoms.⁹¹ This is because the binding of the ligand can increase the volume of the active site by 80%. Additionally a structure where two molecules of ketoconazole were found bound together,⁹² raised the issue of multiple simultaneous binders. This ability in accommodating multiple binders is of further concern on the side of the enzyme promiscuity since, shape, and charge distribution of the active site can be modulated by a second ligand.⁹²

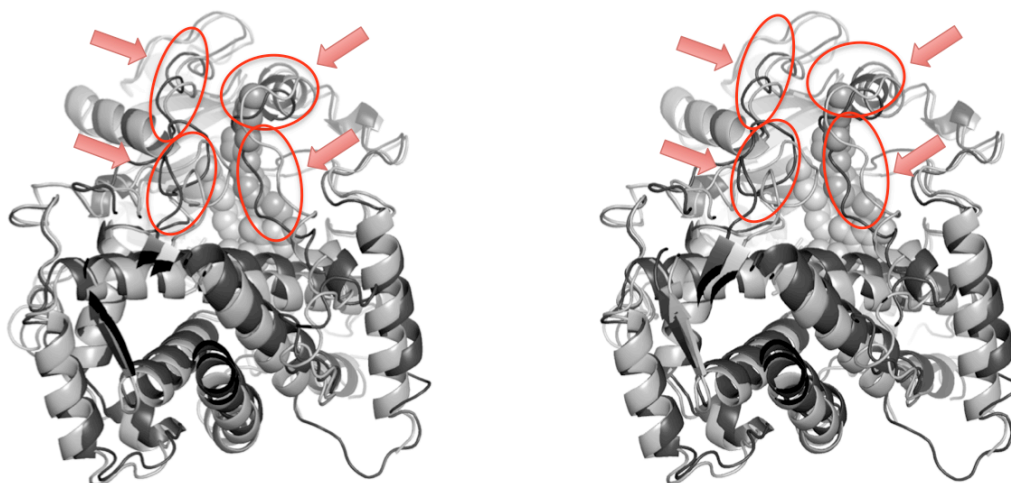


Figure 7: Superposition of CYP3A4 structures (3D image) in the apo form (black, PDB code: 1TQN) and in the holo-form (gray, PDB code: 2V0M). Highlighted (red lines) are the evident conformational changes over binding of the two ketoconazole molecules (gray spheres).

Observations on rabbit CYP2B4 by isothermal calorimetry (ITC) and x-ray crystallography demonstrated how different ligands induce different conformational changes in the same protein.⁹³⁻⁹⁵

The presence of many different protein conformations, active-site rearrangements, and the possibility of simultaneous multiple binders are possibly explaining kinetic data from ligands displaying non-Michaelis-Menten kinetics and give further insights to evaluate possible drug–drug interactions.⁹²

When considering CYPs as targets, such substantial conformational changes due to ligand binding must be considered. Therefore in molecular docking applications, a flexible docking protocol is the most appropriate.

2.2 Approaches in cytochrome modeling: 3D-QSAR

Different computational approaches have been applied to the modeling of cytochromes. Among others, 3D-QSAR (three dimensional quantitative structure activity relationships) is an approach that can be used to predict binding affinities.

3D-QSAR methods, look for mathematical correlations between biological activities and properties (descriptors or quasi-atomistic models in the case of the Quasar technology) derived from 3D molecular structures (e.g. atom position, physico chemical properties) by means of statistical methods.⁹⁶ A biological response can be considered as an interaction in 3D space of two entities. Therefore, those entities and the related interactions can be characterized by a 3D distribution of properties either being ligand based, or protein based, or both.

The above-mentioned mathematical correlation is usually function of the superposition of ligands based on pharmacophore information, or on their bioactive pose. The main problem related to the obtainment of these superpositions (and alignments) is the minimization of the bias.

3D alignments are predominantly evaluated by means of their similarity in terms of energy fields. One of these methods is CoMFA (Comparative Molecular Field Analysis).⁹⁷ This methodology is based on the superposition of compounds and on the calculation of their steric and electronic properties on a 3D grid, which is created around the superimposed ligands. The statistical model obtained through partial least squares (PLS) fitting, is built based on the relationships

between interaction fields and the affinity data. CoMFA is one of the techniques often used to generate 3D-QSAR models at the CYPs.^{81, 98}

Most of the 3D-QSAR models are generated using linear statistical regression methods. These methods include multiple linear regression (MLR), principal component analysis (PCA), principal component regression (PCR), and partial least squares (PLS).⁹⁹

When establishing 3D-QSAR models or more generally when modeling CYPs structures and ligand binding to these enzymes, one of the main issues to remind is that the quality of the model will be directly dependent to the quality of the alignment as well as on the quality of the properties/descriptors, which are calculated or extrapolated from these alignments.

In a QSAR model, the number of descriptors adopted should not be larger than five or six (generally the proportion is 1:10, one descriptor for ten instances), for the complexity of their interpretation, and because following the “principle of parsimony”: ‘One should not increase beyond what is necessary (the number of entities required) to explain anything’.¹⁰⁰ A too high number of descriptors, a too small size of the training set, along with deletion of outliers, or the use of certain statistical techniques can bring to the over-fitting of the data¹⁰⁰ and therefore to a non-predictive model.

QSAR models must be evaluated based on their level of predictivity and on their fit. The fit can be evaluated through their coefficient of determination (r^2) or through the F-test of the regression. The predictivity is generally assessed by cross validation in groups (leave-many-out) or leave-one-out cross validation (LOO). In the latter one high values of q^2 do not necessarily correlate with the test set predictivity.^{101, 102, 103} Further ways to test the effective predictivity of QSAR models are the ones adopted in the study on CYP2D6 and CYP2C9 presented in this thesis: Y-scrambling and the test of completely external dataset of compounds where biological activities are available.

One frequent assumption, done when modeling CYPs behavior, but more generally when modeling the binding of ligands to proteins, is that all the ligands bind in their low-energy active conformation, in a rigid binding mode, and at the same binding site. This can be misleading and not accounting for temporary states which can occur during ligand-protein recognition or even more generally during chemical reactions.¹⁰⁴

2.3 Mixed-model QSAR at cytochrome P450 2D6 and 2C9

In this work, the importance of a correct simulation of the active site flexibility is emphasized comparing atomic mobility from X-ray crystal structure determinations to the average movements within molecular dynamics (MD) simulations. The found correlation is a further justification for adopting a flexible docking protocol.

Affinities (K_i) for binding and non-binding molecules (120 for CYP2D6 and 141 for CYP2C9) were selected from a dataset of 500 drug-like compounds.¹⁰⁵ We adopted a mixed-model protocol where a ligand-based pre-alignment is followed by automated, flexible ligand docking to the enzyme structure. The quantification is finally performed through multi-dimensional QSAR (mQSAR) using Quasar.

Quasar^{106, 107} is a receptor-modeling concept that allows for multi-dimensional QSAR. In *Quasar*, the binding site of the protein is represented by a surrogate, which consists of a three-dimensional surface, surrounding the ligands superimposed in their bioactive conformation (as obtained, for example, from docking studies at the true biological receptor) at van der Waals distance. The topology of this surface mimics the shape of the binding site. This surface is then populated with quasi-atomistic properties, corresponding to those of the amino- acid residues. The 4th dimension is able to deal simultaneously with different conformations, orientations, and protonation states. Some of these approaches provide also the possibility of simulating induced fit in an explicit manner, by means of a topological adaptation of the model of the binding-site surface to the individual ligand molecules. Different induced-fit protocols and solvation models constitute the fifth and sixth QSAR dimension (5D-QSAR, 6D-QSAR),^{107, 108} respectively. Multi-dimensional QSAR (mQSAR) can be considered a supervised method, which generates a model based on a training set to predict values of an external dataset using the produced model.

The models were tested and validated checking the vicinity of the sites of metabolism of the ligands to the iron–heme system, performing Y-scrambling, and supplying external datasets of compounds to the models for CYP2D6 and CYP2C9.

Compounds displaying affinities $\geq 200 \mu\text{M}$ (i.e. considered as weak or non-binders) to the two cytochromes were tested using the same protocol to evaluate the probability of obtaining false positive predictions.

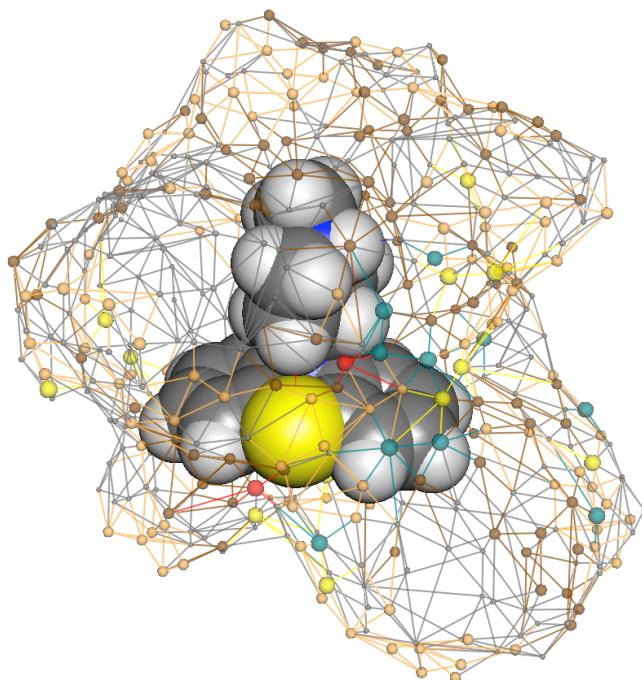


Figure 8: Representation (wireframe and dots) of pseudo-receptor as generated by Quasar for CYP2D6. Highlighted with vdW representation one of the poses of methdilazine.

The validated mQSAR models of CYP2D6 and CYP2C9 were included into the VirtualToxLab,¹⁰⁹ an *in silico* tool for the prediction of the toxic potential (endocrine and metabolic disruption, interference with the hERG ion channel) of drugs and chemicals. The VirtualToxLab currently includes 16 models: 10 nuclear receptors, hERG and five CYP450 enzymes.

3 Computer-aided drug design at the Myelin Associated Glycoprotein (MAG)

3.1 Axonal regeneration in adult mammalian CNS

Axonal regeneration is impeded after injury by inhibitor proteins within myelin and by glial scar formation.^{110, 111} Myelin is actively involved in the inhibition of the growth. Secondly, the scar upregulates inhibitors of regeneration, and it forms a physical barrier to the growth (Figure 9).

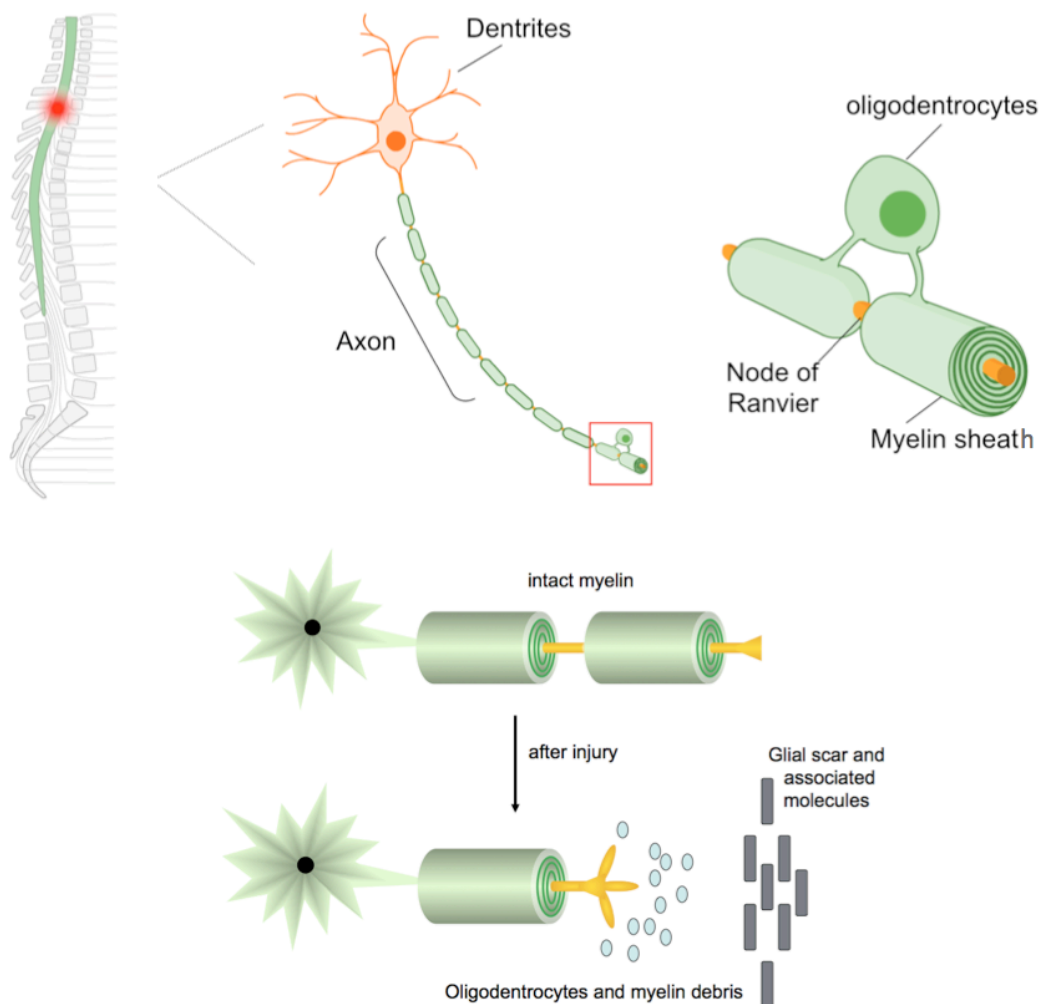


Figure 9: Representation adapted from Spencer et al.¹¹² of spinal chord and blow up of neurite organization with focus on myelin sheath localization and event series occurring after injury.

The adult mammalian central nervous system (CNS) on the contrary of the peripheral nervous system (PNS), does not spontaneously regenerate after injury. The main difference in the capability of regeneration stays in the presence and action of macrophages and Schwann cells which clear myelin after injury in the PNS. Schwann cells continue in their differentiation, downregulating myelin proteins.¹¹³

In the CNS, after injury nerve strands are prevented from re-growing by myelin associated glycoprotein (MAG), oligodendrocyte myelin glycoprotein (OMgp), Nogo receptor (NgR) and p75 neurotrophin receptor (p75^{NTR}) (Figure 10).¹¹⁴

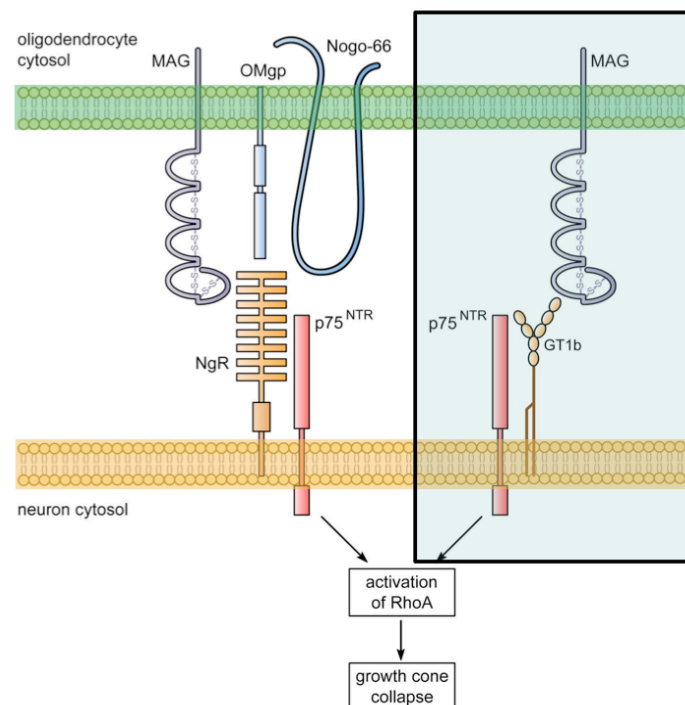


Figure 10: Figure adapted from Filbin et al.¹¹⁴ representing the multiple interactions of glycosyl phosphatidylinositol-linked Nogo receptor (NgR) with MAG, OMgp, and Nogo-66.

A part of this thesis focuses on the Myelin-associated glycoprotein which, was shown to display inhibition of neurite outgrowth.¹¹⁵ MAG, despite the mentioned inhibitory properties, does not inhibit the growth of young neurons, for their ability to regenerate spontaneously *in vivo*.¹¹⁶ This aspect underlines the ambivalent character of this protein.

3.2 MAG: Myelin-Associated Glycoprotein

MAG is a transmembrane glycoprotein belonging to the immunoglobulin (Ig) superfamily. It is selectively localized in the periaxonal Schwann cell and oligodendroglial membranes of the myelin sheaths.¹¹⁷

MAG is a sialic acid-binding protein and it belongs to the SIGLECS family (Sialic acid binding immunoglobuline -like lectins),¹¹⁵ showing homology with its first four Ig-like domains (Siglecs 4a).¹¹⁸ MAG contains 4C-type Ig domains with a single transmembrane domain and a short cytoplasmic tail.

MAG binds preferably to sialoglycoconjugates, in particular to gangliosides¹¹⁸⁻¹²² and a second axonal target is the Nogo-66 receptor (NgR)^{123, 124} to which it binds independently to the sialic acid.

It is known the interaction of MAG with different gangliosides, primarily GD1a (Figure 11), GT1b and GQ1b α (Figure 12).^{114, 125} Gangliosides, sialic acid-containing glycosphingolipids, are highly present in the nervous system especially in the outer side of the membrane. Different ganglioside-binding proteins present structural specificities for their carbohydrate target. When referring to MAG an α 2,3-linked sialic acid moiety is required (Figure 11). The role of gangliosides as MAG ligands along with the role of Nogo receptors and gangliosides have not been yet resolved, but it is supposed that binding to MAG leads to the activation of small kinase RhoA and to growth cone collapse.¹²⁵

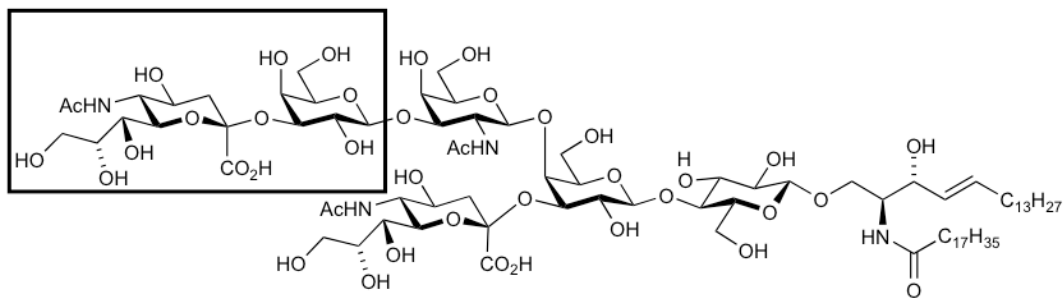


Figure 11: GD1a ganglioside with highlighted the α 2,3-linked sialic acid. Figure adapted from Ernst et al.¹²⁶

It has been shown by Collins et al.¹²⁷ that the Chol-1 family of gangliosides binds with high affinity to MAG. Chol-1 gangliosides carry a α 2,6-linked sialic acid on the GalNAc of a gangliotetraose core (Figure 12).

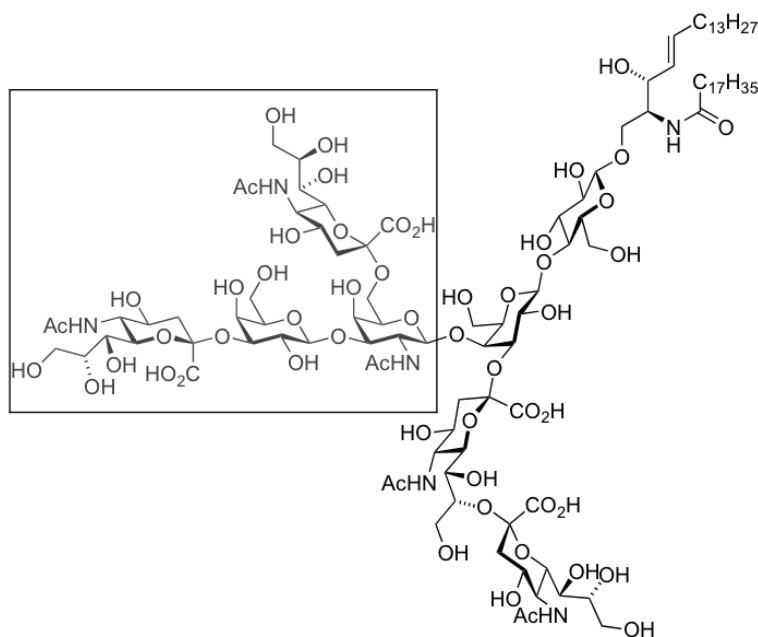


Figure 12: GQ1b α ganglioside with the essential tetrasaccharide binding epitope highlighted. Figure adapted from Ernst et al.¹²⁶

In this study the focus is on the ganglioside-mediated cascade since in certain systems sialidase treatment brought enhanced neural outgrowth. MAG uses sialylated glycans as axonal ligands,¹²⁸ therefore the block of MAG with potent glycomimetic antagonists could offer a valuable therapeutic approach to increase the axonal outgrowth.¹²⁸

3.3 Design of MAG antagonists

Structure-activity relationships revealed that the terminal tetrasaccharide epitope of GQ1b α (Figure 12) presents a stronger binding to MAG compared to GD1a or GT1b.¹²⁷ Moreover, the affinity of a substructure of GQ1b α (Figure 13) has been related to the ability of reversing MAG-mediated inhibition of axonal outgrowth.¹²⁹

Previous studies^{130, 131} revealed that the Neu5c α (2–3)Gal structure, where Neu5 stands for α (2→3)-linked N-acetylneuraminic acid, is not the only part essential for binding. STD NMR and trNOE NMR studies^{132, 133} highlighted that the core Gal β (1–3)GalNAc contributes through hydrophobic interactions to the binding.

The presence of this hydrophobic interaction is further confirmed by its replacement with non-carbohydrate linkers.¹³⁴

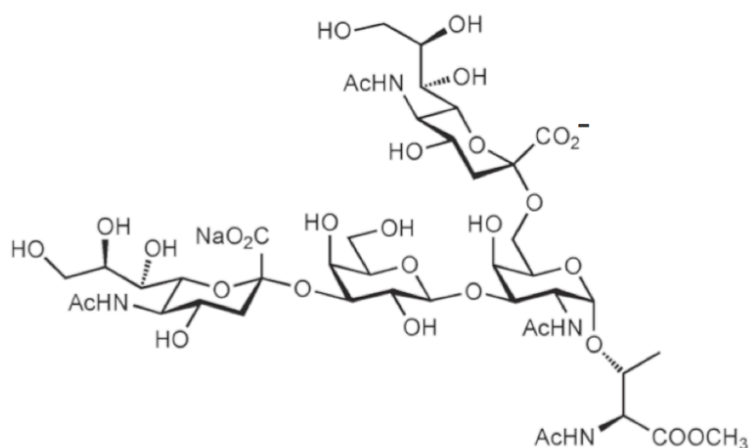


Figure 13: Tetrasaccharide derivative, part of the GQ1b α ganglioside. Figure adapted from Ernst et al.¹²⁶

The substructure of GQ1b α depicted in Figure 13 was investigated by STD-NMR experiments¹³³ and its IC₅₀, was determined in a competitive binding assay to be 300nM.¹³⁵ These experiments highlighted two salt bridges by the two carboxylates of the Neu5Ac moiety and an hydrogen bond established by the C(9)-OH of the α (2-3)-linked Neu5Ac.¹³⁶ Moreover the substructure presents lipophilic interactions of the glycerol side chain of the α (2-3)-linked Neu5Ac, the β -face of the galactose moiety and the N-acetates of the Neu5Ac residues (Figure 14).

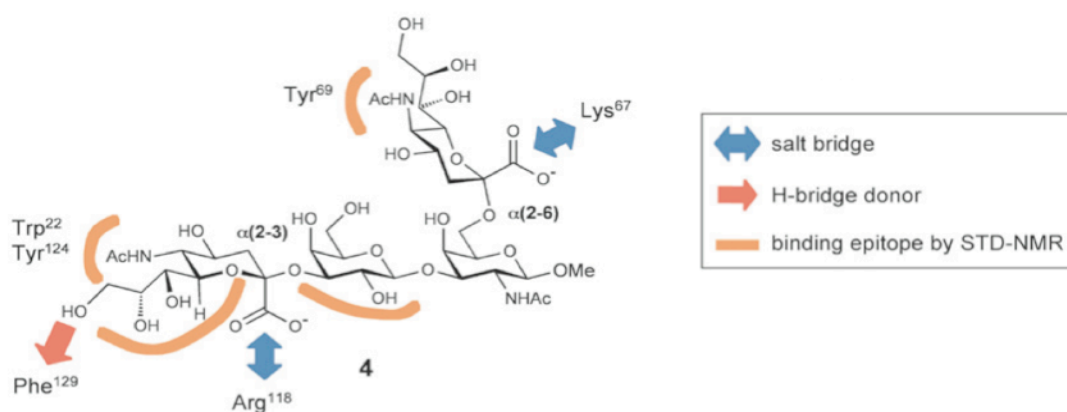


Figure 14: Tetrasaccharide and correspondent interactions to MAG as dedected by NMR experiments.

Attempts to simplify the sialic acid moiety (Figure 15, next page) were ventured by Kelm and Brossmer, which recorded improved antagonist activity when derivatives were modified in position 2- (e.g. Figure 15a),¹³⁷ 5- (e.g. Figure 15b),¹³⁷ and 9- (e.g. Figure 15c).^{137, 138}

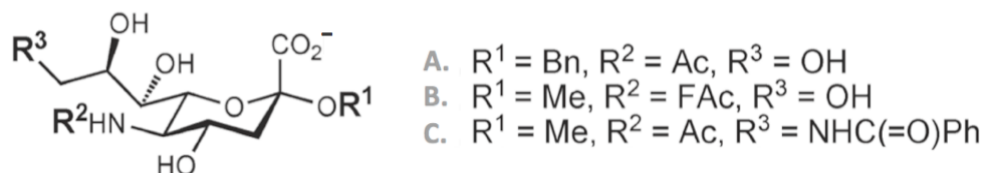


Figure 15: Sialic acid derivatives with different substitutions (A, B, C) in position R1, R2, and R3.

In this thesis, libraries of MAG antagonists are presented along with their binding properties evaluated by hapten binding assay, surface plasmon resonance (SPR), Isothermal Calorimetry (ITC), and competitive NMR experiments. The binding data were rationalized by docking studies, MD and FEP simulations on a homology model of MAG.¹³¹

Methods

1 Water in molecular simulations and drug design

1.1 Cambridge Structural Database search

Hydrogen-bond distances, angles and torsions between water and small molecule crystal structures were collected from the CSD v5.31 (November 2009 + 3 updates). The queries (cf. results session) were generated in ConQuest v1.12 and the histograms in the supplementary material were plotted using Vista v2.1, included in the CSD software suite. The following filters were adopted in the search of crystal structures of small molecules in order to avoid errors or geometries influenced by ions, or calculated coordinates:

- 3D coordinates determined (no calculated atomic positions)
- crystallographic R factor ≤ 0.05
- no disordered, or polymeric structures, no ions present
- only pure organic compounds (no organometallics)

Terminal hydrogen positions were normalized during the search. The hits were defined as structures matching the query search. Multiple fragments (not symmetry-related – for example two carboxylates interacting with water molecule(s) in the same crystal structure) for each hit are possible. A C++ program was devised for transforming bond distances, bond angles, and torsion angles into Cartesian coordinates and to finally produce 3D plots of water oxygen and hydrogen positions (Table 8 and Table 9 – Results).

1.2 *Ab initio* calculation

Preferred interaction geometries and the associated energies of generic functional groups of the ligand with a single water molecule were obtained from *ab initio* calculations using Gaussian 03.¹³⁹ The molecular complexes as well as the isolated entities were fully optimized in the gas phase using the Møller-

Plesset perturbation theory with 6-311++G(d,p) basis set.^{140, 141} At the optimized geometries, an analysis of the vibrational modes was performed to confirm that they represent true minima (without imaginary frequencies detected). The final interaction energy, ΔE , was calculated as the difference between the energy of the complex and the sum of energies of its respective isolated entities. Finally, ΔE was corrected for the superposition error of the basis set using the counterpoise method.^{142, 143} The generic functional groups were then ranked according to their hydration propensity based on the interaction energy (cf. Figure 25).

1.3 Organization of the algorithm

Using a concept similar to the solvation module of Yeti,¹⁴⁴ vectors originating from hydrogen-bond donors (marking the ideal position of a hydrogen bond acceptor, hydrogen extension vectors, HEVs) and hydrogen-bond acceptors (marking the ideal position of a hydrogen-bond donor, lone pair vectors, LPVs) are calculated to identify ideal positions for water molecules. These vectors are generated using specific geometries for each functional group based on data retrieved from the CSD search (cf. Table 8, 9).

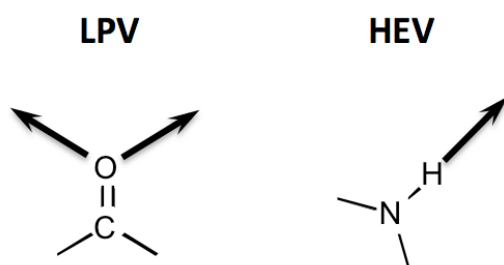


Figure 16: Example of LPV and HEV vectors for carbonyl and amine functional groups.

Halogen atoms present both an electropositive region (also called “corona”) along the extension of the C–X (X=Cl, Br, I) bond and an electronegative region along the axis perpendicular to these bonds (also called “belt”), displaying an amphoteric character in ligand–protein interactions (Figure 17).¹⁴⁵ Moreover, intermolecular distances increase together with the increased polarizability of heavy halogen atom involved in the interaction.¹⁴⁵

In AcquaAlta, apolar hydrogen atoms are considered as potential hydrogen-bond donors¹⁴⁶ and halogen atoms as hydrogen-bond acceptors. For both the cases, in order to identify the starting position of the search for an interaction counterpart, the vectors are considered as the extensions of the carbon-hydrogen and carbon-halogen bond, respectively.

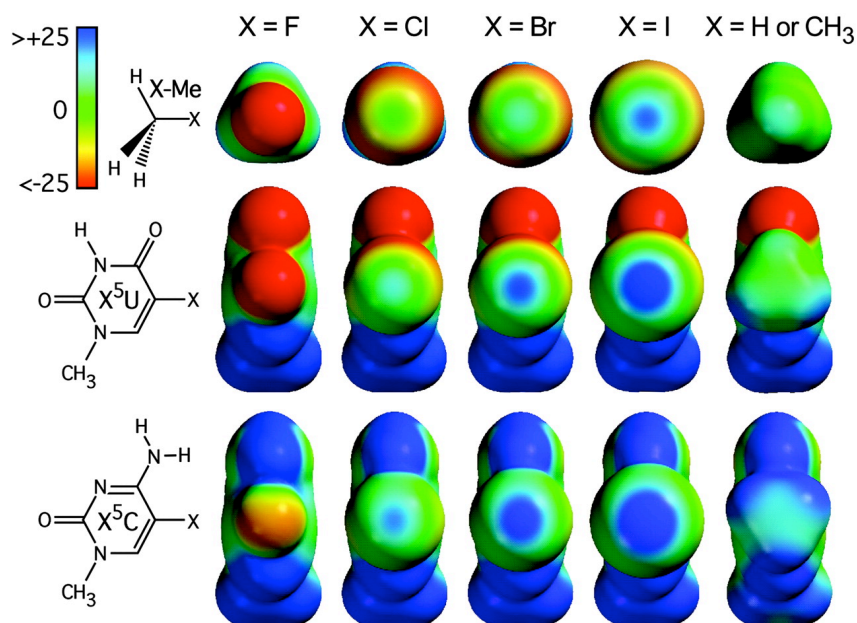


Figure 17: Figure and caption from Auffinger et al.¹⁴⁷ showing *ab initio* electrostatic potential surfaces of halogenated model compounds. Halogenated methane (X–Me, Top), uridine nucleobase (X 5U, Middle), and cytosine nucleobase (X 5C, Bottom) are shown looking into the halogen atoms to compare the induced negative (red), neutral (green), and positive (blue) electrostatic potentials around the halogen surfaces. The compounds are ordered (from left to right) from least to most polarizable (F < Cl < Br < I), with the last column showing, for comparison, the potential surface of methane, methylated uridine, and methylated cytosine.

Figure 18 (next page), depicts the underlying organization of AcquaAlta. First, HEV and LPV vectors are generated (Figure 18A), then, those vectors associated with atoms already engaged in inter- and intramolecular hydrogen bonds are marked as “saturated” and are not further considered (Figure 18B). Distances between unsaturated vectors are calculated and considered as potential water sites if the following criteria are met: distances of LPV–LPV and LPV–HEV below 2.8 Å, which refer to a O–H...O hydrogen-bond length, and HEV–HEV distance below 1.4 Å (Figure 18B). To ensure that an appropriate free volume for a potential water molecule is available, distances from the midpoint of the two vectors tips, are checked against any ligand and neighboring protein atoms. The distance for acceptance, must be larger than the sum of the

individual van der Waals radius of the ligand or protein atom (X) plus 0.96 Å, which refers to the O–H bond length. If the space criterion is met, the water molecule is placed at the midpoint of the vector tips (Figure 18C) into an averaged position determined by the two vectors and then oriented towards its partners (Figure 18D).

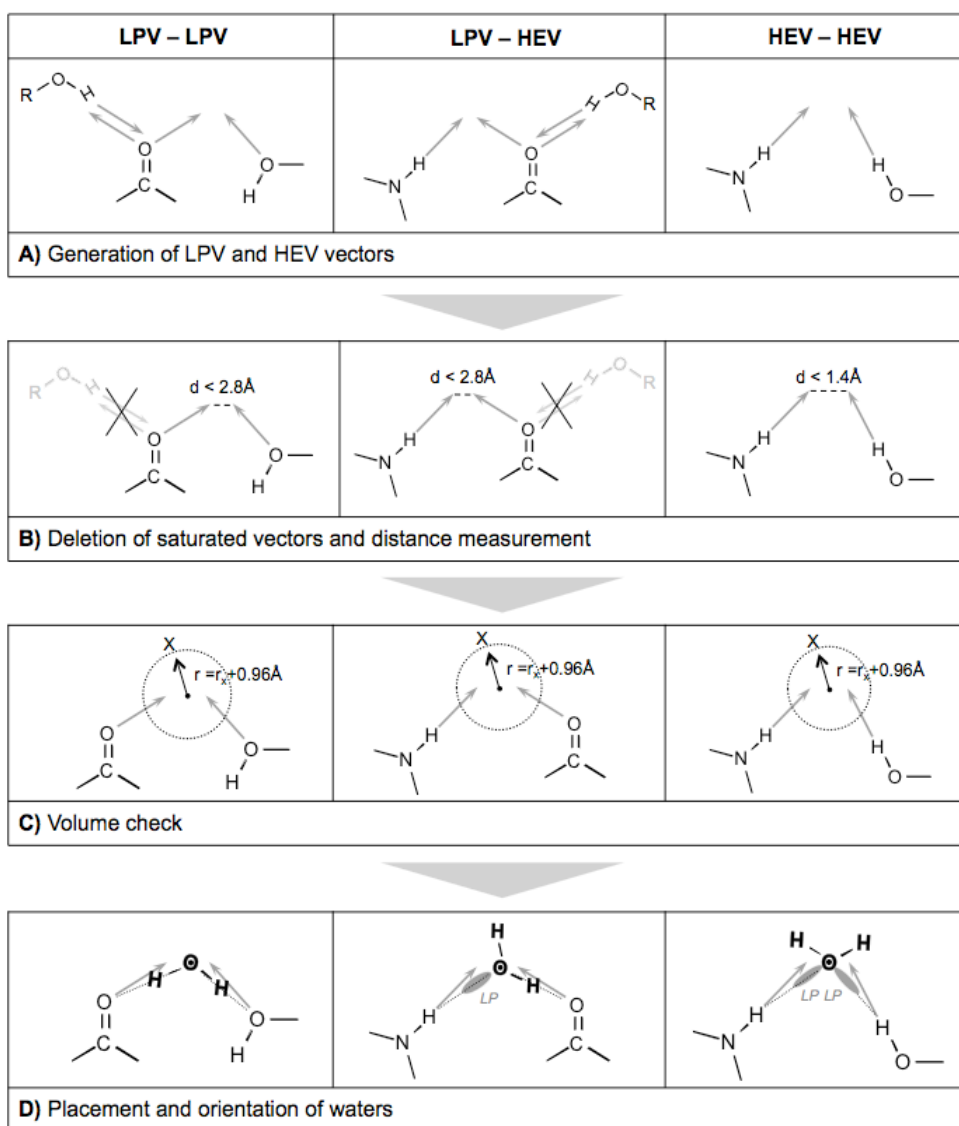


Figure 18: Figure displaying different steps of the vectors generation, scans, placement, and orientation procedure for the possible kinds of vectors combination (LPV–LPV, LPV–HEV, HEV–HEV): A) generation of the LPV and HEV vectors, B) scan for vectors not involved in other inter- or intramolecular hydrogen bonds and measurement of the distance to the unsaturated vectors; C) check of availability of space through distance measurement towards ligand and protein atoms (X) where r_x is the van der Waals radius of the atom; D) placement and orientation of a water molecule towards the atoms origin of the vectors involved in the hydrogen bond.

The detailed organization of the algorithm is shown in Figure 19.

Input structures:

The structures of ligand and protein atoms are supplied in PDB format in order to determine connectivities for all functional groups of ligand and protein, and to generate vectors (HEV, LPV, and originating from halogen and apolar hydrogen atoms). The presence of vectors already engaged in inter- and intramolecular hydrogen bonds (both in ligand and protein) is checked and associated vectors are deleted.

Vector–vector distance check:

Distances are checked for LPV/LPV, LPV/HEV, and HEV/HEV combinations. Distances must remain within the defined thresholds.

Space scan:

Distances towards any ligand and protein atom are calculated to ensure that a sufficiently large volume to accommodate a water molecule is available. To check whether a water molecule has been already placed in the vicinity, distances between water oxygen atoms must be larger than 2.49 Å [sum of van der Waals radii (2.8 Å) multiplied by a factor of 0.891 (threshold value below which the Lennard-Jones 6-12 potential becomes repulsive)].

Type definition:

While generating the positions of bridging water molecules, three different types, with decreasing priority, are defined:

- i) waters bridging interactions within two ligand atoms,
- ii) waters interacting with both ligand and protein,
- iii) waters establishing only a single hydrogen bond with the ligand.

Water placement and orientation:

Finally, the water molecule is placed and oriented towards the origin of the closest unsaturated vector. Vectors are considered based on their hydration propensity as ranked from the *ab initio* calculations (Figure 25 – Results). The geometric criteria for acceptance are loosened from very stringent to moderate giving priority to distance, linearity, and deviation from the lone-pair direction.

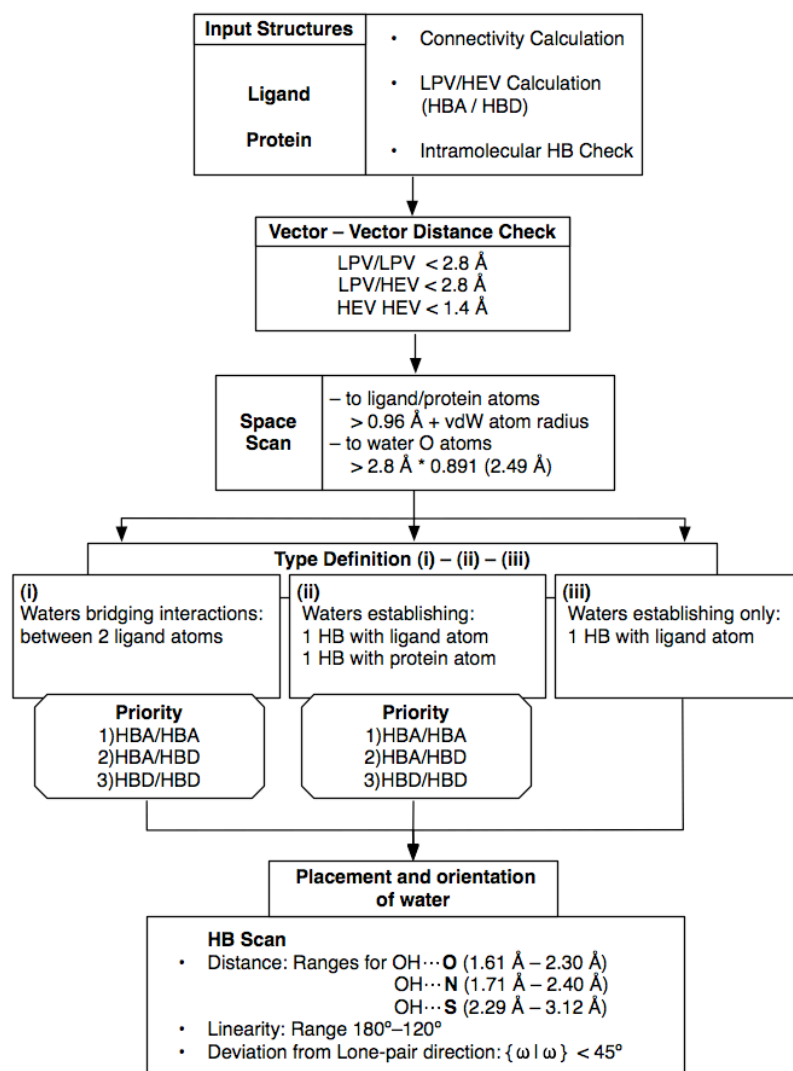


Figure 19: Flowchart of the algorithm organization. Calculations and procedures are performed in the same order as executed the top to the bottom and from left to right. HBA stands for hydrogen bond acceptors and HBD stands for hydrogen bond donors. Criteria for hydrogen bond (HB) acceptance are included.

1.4 Validation of AcquaAlta

Table 6 lists the proteins retrieved from the Protein Data Bank (April 2010) for the purpose of algorithm validation. The crystal structures were selected according to their resolution, the number of bridging waters or because they

were subject of previously published studies dealing with water bridging polar interactions.

The individual PDB files were splitted into protein and ligand part with atom types based on the AMBER* force-field using Yeti¹⁴⁸ and atomic partial charges from the AMBER databased¹⁴⁹. Protonation and tautomeric states for ligand and protein structures were determined using Epik from the Schrodinger software suite.¹⁵⁰

The criteria defining a bridging water molecule were arbitrarily set to a distance within 3.3 Å from any ligand atom and interaction with any kind of polar protein atom within the threshold 2.7–3.3 Å. An energy filter — a water molecule is considered “bridging”, if its total interaction energy, E_{tot} ($E_{ele}+E_{vdW}+E_{HB}$) was equal or lower than -1.0 kcal/mol after rotational optimization (oxygen atoms kept fixed) based on the Yeti force field¹⁴⁸ — was applied in addition.

For the comparison of experimental and calculated waters, the distance threshold for oxygen-water matching was defined as 1.4 Å (50% of a O-H···O hydrogen bond).

Table 6: List of the 20 X-ray structures and relative number of waters, used for the validation of AcquaAlta.

#	Biological Target	PDB code	Res. [Å]	R-factor	# waters
1	Trypsin	2ayw	0.97	0.138	8
2	Dihydrofolate Reductase	3dfr	1.70	0.152	23
3	Thymidin Kinase	1e2k	1.70	0.209	5
4	VEGFR2	1ywn	1.71	0.206	2
5	Glycogen Phosporylase	1a8i	1.78	0.182	13
6	Human Phospodiesterase	1xp0	1.79	0.194	1
7	Beta Trypsin	1bjv	1.80	0.171	2
8	Holo-Glyceraldehyde 3P Dehydrogenase	1gd1	1.80	0.177	15
9	Hsp90	1uy6	1.90	0.184	4
10	AmpC Beta-Lactamase	1xgj	1.97	0.168	2
11	2CDK2	2b53	2.00	0.223	3
12	ACE	1o86	2.00	0.180	6
13	COMT	1h1d	2.00	0.174	4
14	HIV-1 Protease	1hpx	2.00	0.170	4
15	Non-Nucleoside Adenosine Deaminase	1ndw	2.00	0.206	6
16	ACK1	1u4d	2.10	0.205	5
17	Coagulation Factor XA	1f0r	2.10	0.216	2
18	Thymidin Kinase	1kim	2.14	0.209	3
19	EGFR	1xkk	2.40	0.209	2
20	EGFR	1m17	2.60	0.251	1

1.5 Application of AcquaAlta: Docking with Alignator

AcquaAlta was used to re-solvate protein-ligand complexes, previously stripped of the solvent and ligand poses produced by Alignator (14 OppA complexes – Table 7). Finally, the water molecules were optimized using Yeti. The deviation between crystal waters and calculated waters was monitored for each oligopeptide–protein complex using the same distance criteria as in the validation step. (cf. above)

The software AcquaAlta is applied to ligand poses in protein complexes as generated by Alignator.¹⁵¹ This pharmacophore-based alignment tool takes advantage of a conformer pool generated by a conformational search. In our case, we used Macromodel¹⁵⁰, in implicit water, using the OPLS2005 force-field, 20,000 iterations, 1,000 steps for each rotatable bond, energy window of 100 kcal/mol. Each accepted conformer was then aligned to a template molecule (i.e. Lys-Asn-Lys for OppA) based on the matching pharmacophores. Of all possible solutions, only those having maximum number of superimposable pharmacophores were retained. Further information on the approach together with its application are published elsewhere.¹⁵²

Table 7: List of the OppA crystal structures and relative number of waters, used in the automated docking.

#	PDB code	Resolution [Å]	R-factor (work)	Oligo-peptide sequence	# waters
1	1JET	1.2	0.229	KAK	7
2	1JEU	1.25	0.224	KEK	9
3	1JEV	1.30	0.203	KWK	6
4	1B4Z	1.75	0.179	KDK	10
5	1B5I	1.9	0.182	KNK	7
6	1B32	1.75	0.182	KMK	7
7	1B3F	1.8	0.177	KHK	7
8	1B46	1.8	0.177	KPK	6
9	1B51	1.8	0.179	KSK	9
10	1B58	1.8	0.179	KYK	7
11	1B5J	1.8	0.182	KQK	10
12	1B9J	1.8	0.179	KLK	6
13	1QKA	1.8	0.179	KRK	6
14	1QKB	1.8	0.181	KVK	6

2 MeTOX: Metabolism and TOXicology prediction

2.1 Protein structures

For CYP2D6, the only available 3D structure (of the human apoenzyme; PDB code: 2F9Q, 3.0 Å resolution, $R_{free} = 0.286$) was employed.¹⁵³ For the simulations, the mutations that were used to facilitate crystallization were reset to those of the wild type, i.e. D230 back to L230, R231 back to L231, and M374 back to V374.

For CYP2C9, the crystal structure in complex with flurbiprofen (PDB code: 1R9O, 2.0 Å resolution, $R_{free} = 0.236$) was used.¹⁵⁴ The residues (V38-I39-G40-N41-I42 and Q214-I215-C216-C217-N218-F219-S220) not resolved in the electron-density map of the X-ray structures were rebuilt using the PrGen software.¹⁵⁵

Prior to the molecular-dynamics sampling with Desmond,¹⁵⁶ we applied a relaxation protocol divided into: (a) molecular-mechanics minimization with restrained and unrestrained solute, (b) two MD simulations of 12 ps at a temperature of 10 K in the NVT^[1] and NPT^[2] ensemble with non-hydrogen solute atoms restrained, (c) two simulations of 24 ps in the NPT ensemble at 300 K, 1 atm with restrained and unrestrained water oxygen atoms. Two MD simulations of 1.2 ns were then performed using CYP2D6 in complex with methdilazine and CYP2C9 with the (co-crystallized) ligand flurbiprofen.¹⁵⁴

The enzymes were simulated within cubic boxes with a volume of around 60×100×100Å in size, using the TIP3P water solvation model^[3], periodic-boundary conditions and a salt concentration (NaCl) of 0.15 M, mimicking physiological conditions.

[1] *NVT or Canonical ensemble*: During the MD simulation the parameters of moles (N), volume (V) and temperature (T) are considered conserved.

[2] *NPT or Isothermal-Isobaric ensemble*: moles (N), pressure (P) and temperature (T) are considered conserved.

[3] TIP3P water model: Water molecules are represented through a three-interactions site model (i.e. for each atom of the molecule is assigned a point charge).

In order to analyze mobility of the C α atoms, the thermal displacement factor (B_{iso} ^[4]), as obtained from the X-ray structure determination was correlated to the motion observed in our MD simulations (*rms* deviations analyzed at 1.2 ps intervals). Alignment, fit, and *rms* deviations were calculated for each frame of the simulations relative to the starting frame of the trajectory using the McLachlan algorithm¹⁵⁷ as implemented in the program ProFit (Martin, A.C.R., www.bioinf.org.uk/software/profit). Next, a trajectory-cluster analysis using the Hierarchical Cluster Linkage Method was used to identify the most populated clusters of structures. From these clusters, the structure displaying the lowest energy was fully refined in aqueous solution using the OPLS 2005 force field,¹⁵⁸ as implemented in the software MacroModel.¹⁵⁰ The atomic partial-charge model for all ligands (CM1) was generated with AMSOL;¹⁵⁹ partial charges for the proteins were assigned from the AMBER database.¹⁴⁹

The iron–heme system was modeled in the oxenoid iron (formally [Fe⁵⁺... O²⁻]) as in a previous study for CYP3A4.¹⁶⁰ For a realistic simulation of the system was used a directional force field (Figure 20), which also features a metal function allowing for charge transfer (particularly among oxygen, iron, sulphur, and the heme nitrogen atoms) and for functions controlling symmetry, directionality and ligand-field stabilization (LFSE).¹⁴⁸ In particular, the dynamic charge-transfer function enables for a smooth and continuous transition between ionic and covalent character of the metal-ligand bond.

$$\begin{aligned}
 E_{total} = & \sum_{bonds} K_r (r - r_{eq})^2 + \sum_{angles} K_\theta (\theta - \theta_{eq})^2 + \sum_{torsions} \frac{V_n}{2} [1 + \cos(n\phi - \gamma)] + \\
 & \sum_{nb\ pairs} \frac{q_i \cdot q_j}{4\pi\epsilon_0 D(r) r_{ij}} + \sum_{nb\ pairs} \left(\frac{A}{r_{ij}^{12}} - \frac{B}{r_{ij}^6} \right) + \\
 & \sum_{H\ bonds} \left(\frac{C}{r_{ij}^{12}} - \frac{D}{r_{ij}^{10}} \right) \cdot \cos^2(\theta_{Don-H...Acc}) \cdot \cos^n(\omega_{H...Acc-LP}) + \\
 & \sum_{metal\ pairs} \frac{q_i^{CT} \cdot q_j^{CT}}{4\pi\epsilon_0 D(r) r_{ij}} + \sum_{metal\ pairs} \left(\frac{E}{r_{ij}^{12}} - \frac{F}{r_{ij}^{10}} \right) + \\
 & (E_{MC} + E_{LFS}) \cdot \prod_{angles} \cos^2(\Psi_{Lig-Met-Lig'} - \Psi_{eq}) \cdot \frac{1}{n} \sum_{ligands} \cos^n(\omega_{Met...Lig-LP})
 \end{aligned}$$

Figure 20: Yeti force field¹⁴⁸ with metal-function terms highlighted.

[4] B_{iso} – B-factor or thermal displacement factor: It reflects the mean displacement of an atom $\langle u \rangle$, thus the magnitude of oscillation.

2.2 Ligand structures and affinity data

The experimental affinities of 120 compounds (56 binders and 64 non-binders) for CYP2D6 and 132 compounds (76 binders and 56 non-binders) for CYP2C9, were kindly provided by D. R. McMasters¹⁰⁵ (Compound structures, Appendix 2 – Supporting Information). Compounds exhibiting poor solubility, non-Michaelis–Menten kinetics, and displaying fluorescent interference during the competitive inhibition assay used to obtain their affinities were excluded. The selection of inhibition data originates from a unique dataset of 500 drug-like compounds from the Merck sample repository, measured using recombinant enzymes. Therefrom a subset was selected according to the following criteria: $K_i < 200 \mu\text{M}$, compounds associated with a trade name (to identify the active stereoisomer), and ligands whose database identifier or smiles code were unambiguous. While for CYP2D6, the activity range of the employed compounds extended over 3.3 orders of magnitude, the range of CYP2C9 was only 2.4. Therefore, we added nine compound from a different source¹⁶¹ (Table 3, Appendix 2 – Supporting Information), yielding a total of 141 ligands for the CYP2C9 data set. The selection used to estimate the potential of false-positive predictions (64 for CYP2D6 and 56 for CYP2C9) originates from compounds with affinities $\geq 200 \mu\text{M}$ of the released Merck sample repository. The most probable protonation state at pH 7.4 was predicted through the software Epik from the Schrodinger software suite; in case of proton dissociation constants close to pK 7.4, both protonation states were examined.

After generation, the 3D structures of all ligands were subjected to a conformational search in aqueous solution (5,000 rounds of Monte Carlo search) performed with MacroModel from Schrodinger Inc. For each ligand, all conformations within 10 kcal/mol from the lowest-energy entity were retained and composed into a 4D dataset taking in account different conformations, poses, protonation states and tautomers.

2.3 Prealignment and automated, flexible docking

As the crystal structures of both cytochromes do not contain any of the dataset ligands, the template complexes were generated by manual docking. The template molecules are the most affine ligands, identified for the different chemical classes, which were manually docked into the active sites based on

site-directed mutagenesis information¹⁶²⁻¹⁶⁶ as well as based on the vicinity of sites of metabolism towards the oxygen-iron system of the heme.

The docking algorithm uses random orientations of the global minimum conformer as a starting point. In order to facilitate the automatic search for reasonable binding poses, besides these random orientations a pharmacophore-based alignment tool (Alignator¹⁵¹) was employed. Alignator takes advantage of the conformer pool generated in the abovementioned conformational search. Each accepted conformer is aligned to a template molecule (for CYP2D6 a single and for CYP2C9 four templates were used) based on the matching pharmacophores. Of all possible solutions, only those having maximum number of superimposable pharmacophores are retained. Alignment solutions leading to ligand bumps with unmovable parts of the protein are filtered out (in this concept, only locally induced fit related to the side-chain movement is supported: heme, backbone and C β atoms, as well as bridged cysteine and proline residues are considered rigid).

In the next step, the residues in the active site (having flexible side-chain atoms within 6.0 Å from the ligand) are subject to conformational analyses. For each side-chain, only sterically allowed low energy conformations are saved. A comprehensive combinatorial algorithm is employed in order to identify the best arrangement of all side-chains having minimal number of bumps and the best H-bond interaction energy with the ligand and with each other. Up to four protein-ligand complexes with protein side-chains reorganized for an optimal fit are prepared based on the following criteria (Figure 21, Complex A–D): A) the best *rmsd* with the template; B) the best *rmsd* with the template, without bumps into unmovable protein atoms; C) the best *rmsd* with the template without bumps into unmovable backbone and side-chain atoms and D) the best volume overlap with the template. These selection criteria ensure that both similar as well as rather different compounds are rationally pre-aligned in the active site facilitating the subsequent automatic docking procedure.

To fully simulate local induced fit in Cheetah,^{160, 167} an extended active-site region (any atom of amino acid residue located within 12 Å of any atom of the ligand molecule) is considered flexible. All poses (ligand conformations, orientations, positions) within 25% of the lowest-energy pose (typically 8-15 kcal/mol) featuring a minimal *rmsd* > 1.0 Å are composed in a 4D data set and forwarded to mQSAR for quantification of the binding affinity. The flow-chart of the pre-alignment/docking along with the applied Monte Carlo protocol are shown in Figure 21.

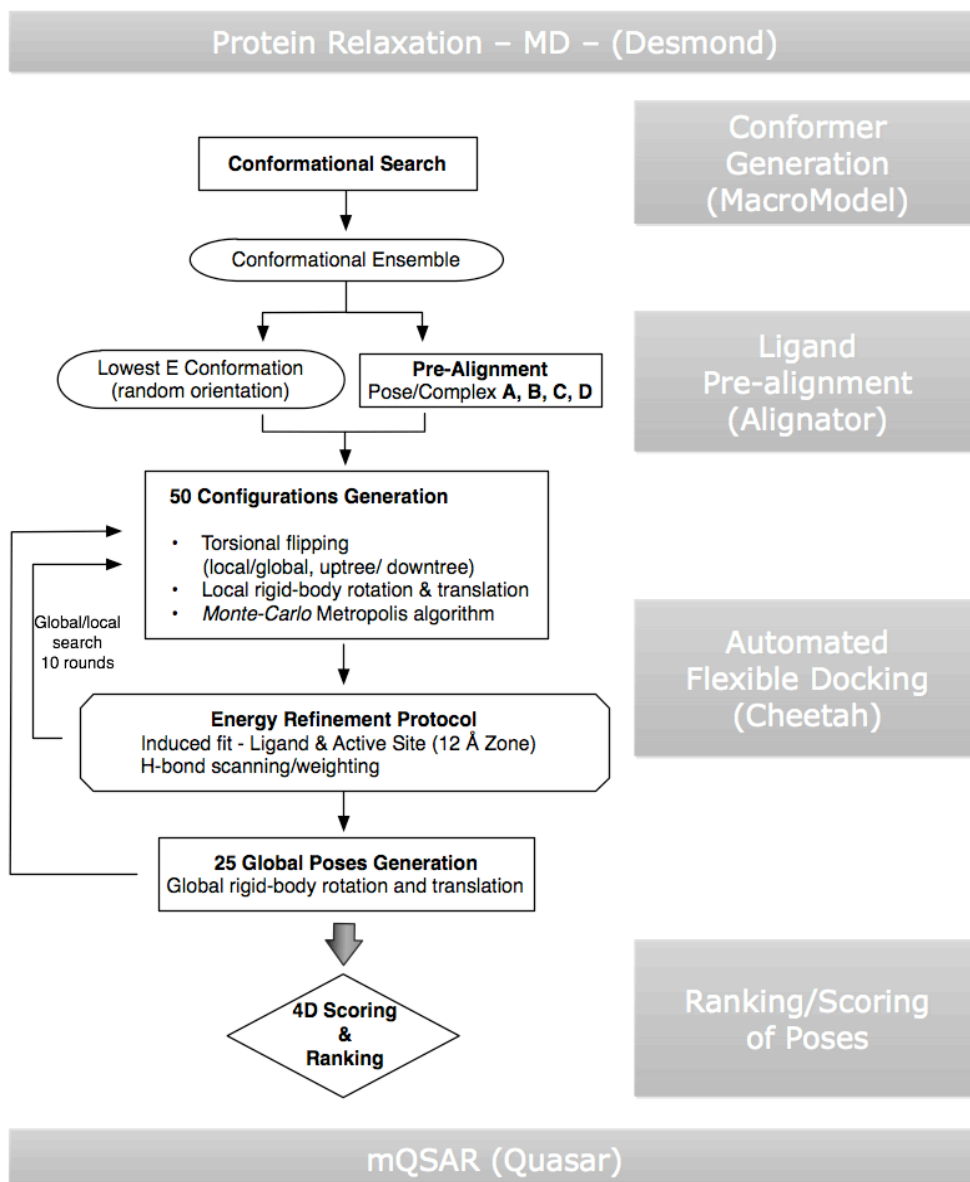


Figure 21: Protein relaxation, pre-alignment and flexible docking protocol. For each ligand, a conformational search generates a pool of conformers of which those within 5 kcal/mol from the lowest-energy conformation found were considered. Each conformer of the ensemble is pre-aligned to a template, thereby allowing side-chain rearrangement within the active site (*software Alignator*). Up to four of the different ligand-protein complexes (Complex A–D) generated are used as starting poses in the automated, flexible docking (*software Cheetah*). All configurations obtained by either the minimized pre-aligned structures or the Monte-Carlo sampling are scored and ranked. (Additional information is provided in Table 4, Supplementary Material).

2.4 Multidimensional QSAR:

The employed mixed-model approach combining automated, flexible docking and multi-dimensional QSAR yields more reliable binding energy predictions than a direct scoring of the ligand-protein interactions of the experimental structure.¹⁶⁸

Quasar is a quasi-atomistic modeling concept that allows for mQSAR. The details of the software are published elsewhere,^{106, 108} and are, therefore, only briefly summarized here. In Quasar, the binding site of the protein is represented by a surrogate, which consists of a three-dimensional surface at van der Waals distance, surrounding the ligands superimposed in their bioactive conformation (as obtained, for example, from docking studies at the true biological receptor). The topology of this surface mimics the shape of the active site. This surface is then populated with quasi-atomistic properties, taking into account the potential of an amino acid to form a salt bridge and to act as hydrogen-bond donor and/or acceptor. Hydrogen-bond flip-flop, solvent and for contributions from neutral, positively and negatively charged and hydrophobic residues are also accounted. In addition to accepting 4D compound input (conformations, poses, protonation states, tautomers), Quasar allows the simulation of induced fit (corresponding to side-chain flexibility and moderate backbone motion at the true biological receptor), whereby six different protocols are evaluated simultaneously (5D-QSAR). The model family, typically consisting of 200 models, is evolved using a genetic algorithm and provides an averaged prediction for each compound along with the variation over the model family. Quasar employs the following scoring function derived from the directional Yeti force field¹⁴⁸ (eqn 1):

$$E_{\text{binding}} = E_{\text{ligand-receptor}} - E_{\text{ligand desolvation}} - E_{\text{ligand internal strain}} - T\Delta S - E_{\text{induced fit}} \quad (\text{Eqn 1})$$

$$\text{where } E_{\text{ligand-receptor}} = E_{\text{electrostatic}} + E_{\text{van der Waals}} + E_{\text{hydrogen bonding}} + E_{\text{polarization}}$$

In the simulations, ligands were represented as 4D ensembles. As a selection criteria for the training:test set splitting, the chemical diversity of the compounds was applied yielding a 46:10 (training:test set) ratio for CYP2D6 and 68:17 for CYP2C9, respectively. The compounds were grouped aiming for the maximal diversity in term of chemical scaffold and assigning to the training set the most and the least active compounds. A family of 200 active-site models for each of the two cytochromes was evolved for 180 generations (36,000 crossovers) in the case of CYP2D6 and 40 generations (8,000 crossovers) for CYP2C9.

2.5 Virtual screening applied to toxicology prediction (Virtual Tox Lab)

The VirtualToxLab (VTL) – Figure 22 – is an *in silico tool* aiming to generate alerts and predict the toxic potential (endocrine and metabolic disruption, interference with the hERG ion channel) of drugs, chemicals and natural products. It simulates and quantifies their interaction towards androgen, aryl hydrocarbon, estrogen α/β , hERG, glucocorticoid, liver X, mineralocorticoid thyroid α/β , and peroxisome proliferator-activated receptor γ as well as for the enzymes CYP450 1A2, 2A13, 2C9, 2D6 and 3A4. These molecular targets are known or suspected to trigger adverse effects.

The technology produces binding affinities based on automated flexible docking combined with multi-dimensional QSAR (mQSAR). VTL allows to visually inspect the predicted binding modes at a molecular level for each of the 16 macromolecular targets, in real time 3D. Figure 22 shows the technical flowchart of the VirtualToxLab.

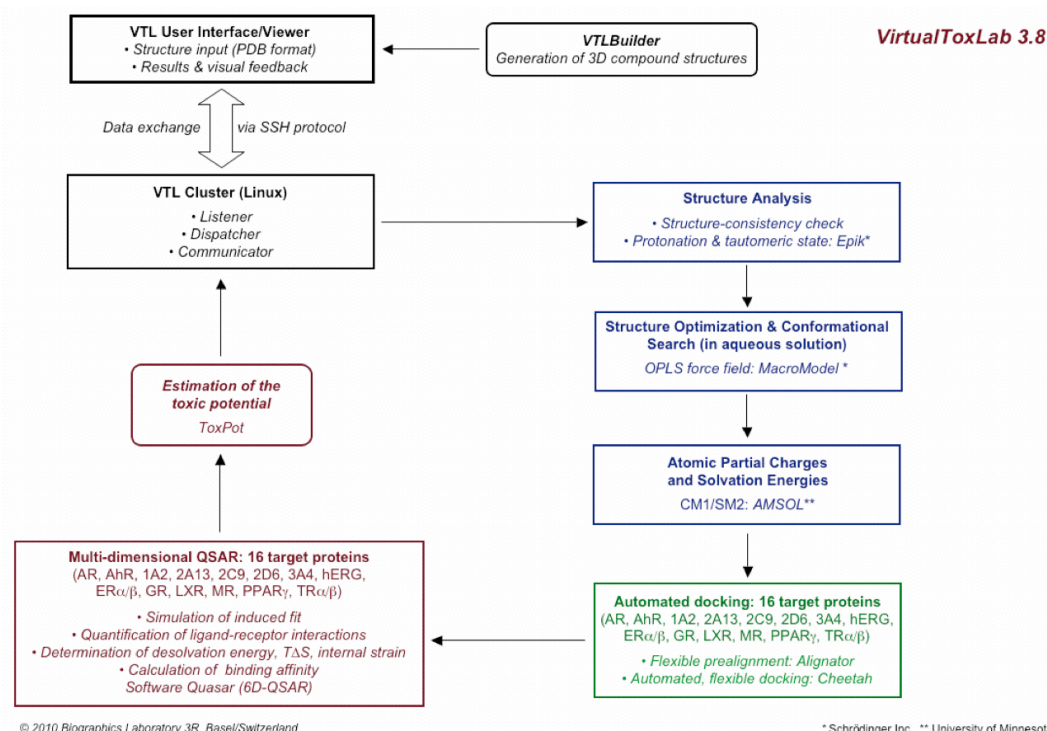


Figure 22: Organization of the VirtualToxLab procedure.

The two validated models of CYP2D6 and CYP2C9 were inserted in VTL and were used to estimate binding affinities and to estimate toxic potentials for all the 2,508 compounds of the *Biograf* Database. The database contains compounds belonging to the following classes:

1. Environmental chemicals
2. Food additives and colorants
3. Antibiotics
4. Psychotropic drugs
5. Blockbusters
6. Withdrawn drugs
7. Designer drugs
8. Various drugs
9. Vitamins
10. Natural compounds
11. Cosmetic chemicals
12. Various chemicals
13. Pesticides
14. Miscellaneous compounds
15. Doping: compounds banned by the National Collegiate Athletic Association (NCAA)

3 Computer-aided drug design at the Myelin Associated Glycoprotein (MAG)

3.1 Docking to the MAG homology model

Docking programs aim to identify low energy poses of a small-molecule ligand in the binding pocket of a protein. Docking simulations are usually implemented in a larger framework comprehending two principal phases: docking and scoring and are classified if accounting for the flexibility either of the ligand or of the protein or of both: (i) rigid docking where both ligand and protein are considered rigid, (ii) semi-flexible docking where ligand is considered flexible and the protein is kept rigid, (iii) and fully-flexible docking where both ligand and protein are considered flexible.

An X-ray structure of MAG is not presently available. A homology model of MAG based on the three-dimensional structure of sialoadhesin was built and previously published by Ernst and coworkers.¹³¹ The model (mouse; UniProt P20917) is based on the 3D structure of sialoadhesin¹⁶⁹ (mouse; Siglec-1, PDB code 1QFP, 2.8 Å resolution) and it was generated using the standard protocol from Fugue¹⁷⁰ and Orchestrar^{171, 172} as implemented in Sybyl 7.3.¹⁷³

The identity in terms of primary sequence is of 29,9% and using multiple sequence alignment with Clustal W¹⁷⁴ the binding site homology was calculated to be 52.5% and 50% for the binding site and the protein, respectively. The model after generation was fully refined in aqueous solution using AMBER 7.0.¹⁷⁵

MAG features a shallow binding site, located on the surface of the protein. Due to the binding site localization, its high solvent accessibility, and protein side-chains with high conformational freedom compared to buried binding sites, the only feasible approach for such investigation was considered allowing for full flexibility. All the inhibitors (cf. Results) were generated using MacroModel and a conformational search was performed to identify low-energy conformations suitable for docking. MacroModel was used for the conformational search of the ligand using the OPLS2005 force field, in aqueous solutions for 5000 iterations.

All the antagonists manually docked to the homology model of MAG, used the salt bridge to Arg 118 as anchoring point, and further interactions were identified for each antagonist. The complexes were minimized using MacroModel.

3.2 Molecular Dynamics

Molecular dynamics simulations is one of the approaches which allow for full flexibility of both ligand and protein. Using MD simulations, it is possible to monitor interactions between ligand and protein in terms both of space and time. MD simulations are based on the Newton's equation (Equation 2) where the force on an atom (i) is related to its mass (m) and its acceleration at a certain time (t).

$$\frac{d^2x_i}{dt^2} = \frac{Fx_i}{m_i} \quad (\text{Eqn. 2})$$

All the MD simulations performed on MAG were preceded by a relaxation protocol divided into: 1) Molecular-mechanics minimization with restrained and unrestrained solute; 2) two MD simulations of 12 ps at a temperature of 10 K in the NVT ensemble (in the NVT or canonical ensemble, the parameters of moles (N), volume (V), and temperature (T) are considered to be conserved) and NPT ensemble (in the NPT or isothermal–isobaric ensemble, moles (N), pressure (P), and temperature (T) are considered to be conserved) with non-hydrogen solute atoms restrained, 3) two simulations of 24 ps in the NPT ensemble at 300 K, 1 atm with restrained and unrestrained water oxygen atoms. Figure 23 displays a diagram of the molecular dynamic protocol.

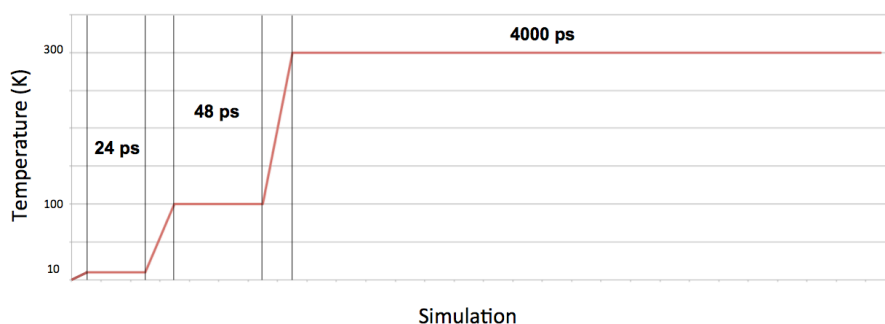


Figure 23: Diagram of the relaxation protocol followed by the MD simulation.

All the different MAG-antagonist complexes were simulated for 4 ns within cubic boxes with a volume of 70 X 100 X 100 Å using the TIP3P water-solvation model [in the TIP3P water model, water molecules are represented through a three-interactions-site model (i.e., for each atom of the molecule is assigned a point charge)], periodic-boundary conditions, and a salt concentration (NaCl) of 0.15 M, mimicking physiological conditions.

3.3 Free energy perturbation (FEP)

The so-called rigorous methods for the estimation of free energies of binding from “not at the equilibrium states” are: free energy perturbation (FEP) and thermodynamic integration (TI). In 1997, Christopher Jarzynski postulated the so-called “Jarzynski equality”,¹⁷⁶ which states that it is possible to relate free energy differences using non-equilibrium processes (e.g. molecular dynamics simulations).

In this thesis, the Free Energy Perturbation (FEP) method^{177, 178} was used, therefore the underlying theory will be explained more in detail.

The estimation of the relative free energy difference (ΔG) refers to the difference between two equilibrium states, which can be referred to as state A and B. Each equilibrium state refers to a ΔG , therefore state A will be identified by ΔG_A and state B will be identified by ΔG_B . The relative free energy difference between the two states will be calculated using the Zwanzig formula¹⁷⁹ (Equation 3) and it will therefore refer to a $\Delta\Delta G$.

$$\Delta\Delta G = \Delta G_B - \Delta G_A = -kT^{-1} \ln [\exp(-\beta\Delta V)]_A \quad (\text{Eqn. 3})$$

In this equation k is the Boltzmann constant and T is the temperature and the term in the square bracket refers to the ensemble average potentials $\Delta V (V_B - V_A)$ as generated by non-equilibrium processes such as molecular dynamics or Monte Carlo (MC) simulations.

Only equilibrium ensembles of configurations in the initial state (not using equilibrium averages) are evaluated. Each configuration is converted to the final state where the difference of energy is evaluated. A Boltzmann average of these energies gives a free energy difference of the two states.

FEP calculations evaluate an ensemble of structures (configurations) assuming that these ensembles have the number of atoms (N) and are performed in the same conditions of pressure and temperature (P, T). These conditions are the so-called isothermal-isobaric ensemble also referred to as NPT ensemble.

In such simulations, the system is gradually transformed from one state to another (from state "A" to state "B"), during which the effect of this perturbation on the free energy is monitored. In practice, this is achieved by defining a morph parameter λ , which is defined to be zero in the initial state (A), and one in the final state (B). By collecting the derivative of the potential energy with respect to the λ during the simulation, and by integrating over λ afterwards, we gain access to the free energy.

One of the many applications for FEP calculations is the prediction of relative binding free energies of protein-ligand complexes. The methodologies applied to the MAG are the following:

- Modification of the ligand, and subsequent estimation of the relative binding free energy of the ligand before and after the mutation.
- Mutation of one protein side chain and subsequent estimation of the relative binding free energy between the two states (before and after protein mutation).

The FEP experiments were performed using the Desmond package,¹⁵⁰ using a sampling time of 1 ns for both ligand and protein mutation. When the ligand mutation is simulated, the FEP is performed for the ligand in complex with the protein and in a solvent box (using the SPC water model). When the protein mutation is simulated, the FEP is performed both in complex with the ligand and in a solvent box (using the SPC water model).

Both FEP and TI are sampling methods that require a high number of configurations to assure that the conformational space of both ligand and protein are correctly evaluated. For this reasons longer simulations are required to obtain more accurate results. This in turn, leads to an increased computational cost for the correct sampling of the two states and the relative convergence of the simulation.

In the case of the MAG system, the 1 ns FEP performed for the protein side-chain mutation used a total of 37 hours of calculations on 48 processors (6 quad-core CPUs with 2.93 GHz clock using hyper-threading technology). For this fact the computational time of each FEP simulation must be carefully considered and accounted.

Results

1 Water in molecular simulations and drug design

1.1 Hydrogen-bond directionality (in water networks)

Based on the CSD searches, detailed information were obtained on the distance, linearity, and deviations from the lone-pair for 16 generic functional groups interacting with water molecules (Figure S1 – Supplementary Information). The preferred geometries were extracted for water molecules interacting with such entities to derive rules to be applied in the solvation algorithm AcquaAlta. The results are listed in Table 8 and 9 for the different functional groups included in the search. The following data are shown for each functional group (entry 1–16): the query and the number of hits and fragments obtained, a 3D representation (side and front view) of water hydrogen atoms (in the case of functional groups acting as HBA) and water oxygen atoms (in the case of functional groups acting as HBD), together with relevant statistical data for distance and angle of interaction.

For this purpose, generic queries were used to compile rules applicable to a wide variety of similar functional groups (e.g. carbonyl oxygen in carboxylic group, amide and ester) and to monitor the general behavior of solvent molecules involved in polar interactions. Quantitative information of specific moieties, (e.g. pyridines vs pyrimidine) and the effects of neighboring atoms of the functional groups are not extractable from the collected data.

The median length of a hydrogen bond is shorter than the sum of the individual van der Waals radii. The difference between strong and weak hydrogen bonds is documented by the following examples. The average mean distance for carbonyl or a phenolic/aliphatic hydroxyl is 2.02 ± 0.25 Å, which is substantially, although not significantly (from a statistical point of view) shorter than the average mean distance for the ones established with ether and nitro groups (2.29 ± 0.29 Å). In addition, the distribution of the water molecules interacting with a carbonyl or a phenolic/aliphatic hydroxyl positions is considerably narrowed

compared to ether or nitro groups. In *AcquaAlta*, the acceptable ranges for the hydrogen-bond distance are shortened compared to the values obtained from the CSD. This allows including charge–atom interactions where the distance is even shorter than the average hydrogen bond distance (the mean distance for $\text{COO}^- \cdots \text{HOH}$ is 1.94 ± 0.27 and the mean distance for $\text{NH}_3^+ \cdots \text{OH}_2$ is 1.86 ± 0.21).

Water prefers to establish hydrogen bonds with a linear $\text{X-H} \cdots \text{Y}$ alignment. Significant deviations of this arrangement are typically caused by additional interactions with nearby atoms.¹⁸⁰ For oxygen acceptor atoms such as carbonyls, hydroxyls, sulfoxides/sulfones, phosphines/phosphones, and nitrogen acceptor atoms as tertiary amines, imines, and nitriles as well as all the hydroxyl and amino donor atoms the median angles vary from 156° to 180° (angles statistics in Table 8). In contrast, for acceptors as e.g. ethers, nitro groups, thioketones, aromatic rings and aromatic and aliphatic “hydrogen-bond donor” groups (Table 4 - entries 15, 16) the median angles substantially deviate from linearity, ranging from 87° (aromatic ring) to 148° (aliphatic CH). In *AcquaAlta*, the $\text{X-H} \cdots \text{Y}$ angle was allowed to be in a range from 120° to 180° , which therefore includes most of the different cases listed above.

Hydrogen bonds are usually directed along the lone pair(s) of the acceptor atom.⁴⁴ Since CSD/Conquest does not allow to define lone pairs, the torsion angle between three atoms of the acceptor group (defining the lone-pair plane) and the water hydrogen atom were analyzed (Figure S1 – Appendix 1 – Supplementary Information).

The 3D plot in Table 8 (entries 1–6, 9–12) shows the distribution of the hydrogen atoms of water molecules around hydrogen-bond acceptors and how water is arranged along the LPVs. Interestingly, water hydrogen atoms in aliphatic and phenolic hydroxyls (entries 2, 3) are positioned also in between the two carbonyl LPVs. Sulfoxide-sulfone and phosphine-phosphone (entries 7, 8) do not show a well-defined distribution because the S-O and P-O bond have only partial double-bond character.⁴⁶ In the latter two cases (entries 7, 8) water hydrogen atoms are distributed equally (i.e. without any spatial preference) forming a “corona”-like pattern around the acceptor oxygen atom. The deviation from the acceptor plane is generally smaller than 30° .⁴⁹ In the algorithm, the optimal positions for interaction with water are considered to be at the end points of LPVs and HEVs. A deviation from the LPV or HEV is allowed up to 45° (Figure 24).

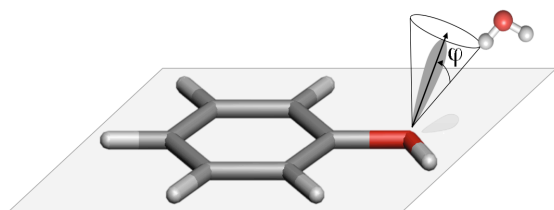


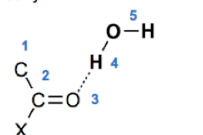
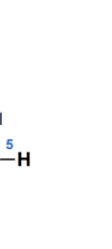
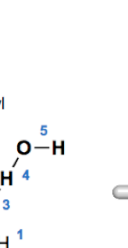


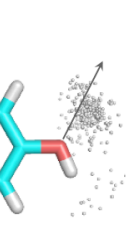


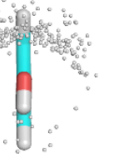
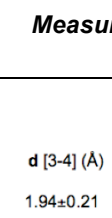
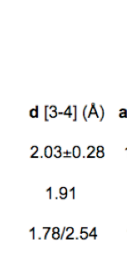
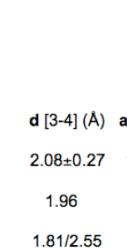
Figure 24: Definition of the allowed deviation (φ) from an ideal LPV (black arrow), in this case assumed at the lone pair position. The same criteria are used for HEVs.

Despite similar studies,^{44, 46-49} no analysis focused exclusively on water molecules interacting with generic functional groups is available. Consistent with a previous study,⁴⁷ this analysis revealed a clear sp^2 lone-pair preference of the carbonyl when interacting with water, but not for the sp^3 lone pair as seen for the ether oxygen atom (entry 4). The results from the CSD searches also confirm a previous study⁴⁸ showing hydrogen bonding preference for heteroaromatic nitrogen acceptors compared to heteroaromatic oxygen acceptors. Our results showed that water hydrogen atoms are less scattered around the lone pair of the nitrogen (entry 6) when compared to the distribution around the lone pairs of oxygen acceptor atom (entry 4).⁴⁸

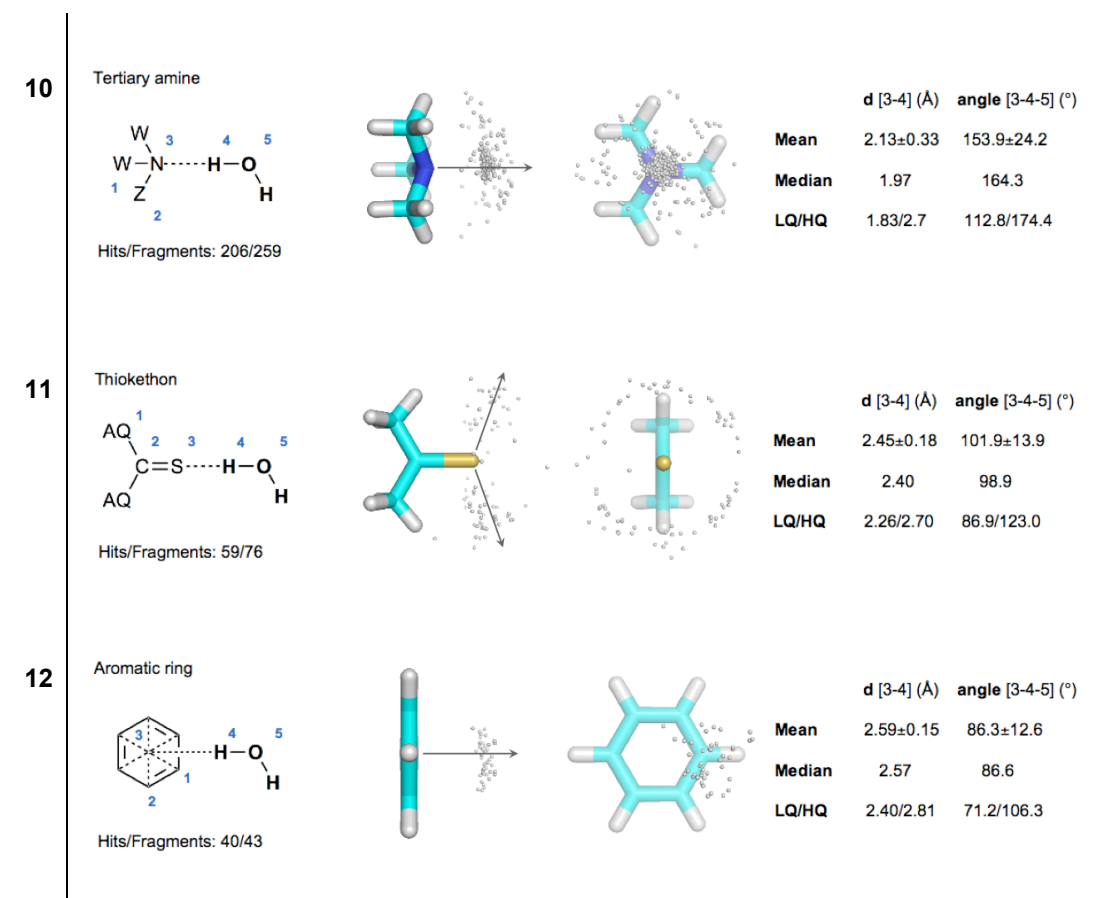
Forty well-defined query hits (entry 12) for the interaction between water (as a HBD) suggest that even fragments generally considered as lipophilic can form favorable interactions with water molecules. In this case, the aromatic ring atoms behave as HBAs and water molecules orient themselves to maximize the interactions of their hydrogen atoms with the π -electron cloud of the aromatic carbons, in a manner comparable with a hydrocarbon σ - π interaction.

Table 8 (Next Page): Summary of the crystallographic data search for functional groups acting as hydrogen bond acceptors (HBA). An entry number indicates each query search. In the left column, the query as submitted to the CSD; the number of hits and the number of fragments examined. In the central column a typical molecule with 3D representation of the positions of water hydrogen atoms (gray spheres), and ideal lone-pair position (gray line). In light-blue sticks general fragments resembling the submitted 2D query. In the right column, statistical results (mean, median, lower and higher quantile) for the values of distance and linearity of hydrogen bonds between functional-groups fragments and water molecules. In the 3D-plots of sulfoxide/sulfone and phosphine/phosphone the queries represent all acceptor atoms to resemble real functional groups. For clarity, in the 3D-plots of the aromatic ring as HBA, the water-hydrogen positions are displayed only for one carbon atom. In the 2D-query representation X stands for any kind of atom, Z for any kind of atom with exception of hydrogen, W for C and H atoms, AQ for C, N, O, S atoms.

HBA: Functional groups acting as hydrogen bond acceptors

#	Query:	Side view	Front view	Measurements
1	<p>Carbonyl</p>  <p>Hits/Fragments: 997/1378</p>			<p>d [3-4] (Å) angle [3-4-5] (°)</p> <p>Mean 1.94±0.21 160.0±17.5</p> <p>Median 1.87 165.8</p> <p>LQ/HQ 1.76/2.24 134.1/175.1</p>
2	<p>Aliphatic hydroxyl</p>  <p>Hits/Fragments: 861/1478</p>			<p>d [3-4] (Å) angle [3-4-5] (°)</p> <p>Mean 2.03±0.28 151.9±23.0</p> <p>Median 1.91 159.7</p> <p>LQ/HQ 1.78/2.54 115.2/173.6</p>
3	<p>Phenolic hydroxyl</p>  <p>Hits/Fragments: 168/267</p>			<p>d [3-4] (Å) angle [3-4-5] (°)</p> <p>Mean 2.08±0.27 148.8±24.2</p> <p>Median 1.96 156.0</p> <p>LQ/HQ 1.81/2.55 112.0/172.7</p>
4	<p>Ether</p>  <p>Hits/Fragments: 375/645</p>			<p>d [3-4] (Å) angle [3-4-5] (°)</p> <p>Mean 2.24±0.31 141.4±25.1</p> <p>Median 2.20 146.4</p> <p>LQ/HQ 1.86/2.66 105.7/171.9</p>

5	<p>Nitro</p> <p>Hits/Fragments: 48/97</p>		<table border="1"> <thead> <tr> <th></th> <th>d [3-4] (Å)</th> <th>angle [3-4-5] (°)</th> </tr> </thead> <tbody> <tr> <td>Mean</td> <td>2.33±0.26</td> <td>112.9±23.3</td> </tr> <tr> <td>Median</td> <td>2.33</td> <td>113.7</td> </tr> <tr> <td>LQ/HQ</td> <td>1.92/2.68</td> <td>78.3/146.1</td> </tr> </tbody> </table>		d [3-4] (Å)	angle [3-4-5] (°)	Mean	2.33±0.26	112.9±23.3	Median	2.33	113.7	LQ/HQ	1.92/2.68	78.3/146.1
	d [3-4] (Å)	angle [3-4-5] (°)													
Mean	2.33±0.26	112.9±23.3													
Median	2.33	113.7													
LQ/HQ	1.92/2.68	78.3/146.1													
6	<p>Imine</p> <p>Hits/Fragments: 451/671</p>		<table border="1"> <thead> <tr> <th></th> <th>d [3-4] (Å)</th> <th>angle [3-4-5] (°)</th> </tr> </thead> <tbody> <tr> <td>Mean</td> <td>2.1±0.30</td> <td>157.9±17.6</td> </tr> <tr> <td>Median</td> <td>1.97</td> <td>164.0</td> </tr> <tr> <td>LQ/HQ</td> <td>1.84/2.64</td> <td>130.6/175.1</td> </tr> </tbody> </table> <p>(**) nitrogen not charged</p>		d [3-4] (Å)	angle [3-4-5] (°)	Mean	2.1±0.30	157.9±17.6	Median	1.97	164.0	LQ/HQ	1.84/2.64	130.6/175.1
	d [3-4] (Å)	angle [3-4-5] (°)													
Mean	2.1±0.30	157.9±17.6													
Median	1.97	164.0													
LQ/HQ	1.84/2.64	130.6/175.1													
7	<p>Sulfoxide /Sulfone</p> <p>Hits/Fragments: 139/271</p>		<table border="1"> <thead> <tr> <th></th> <th>d [3-4] (Å)</th> <th>angle [3-4-5] (°)</th> </tr> </thead> <tbody> <tr> <td>Mean</td> <td>1.96±0.23</td> <td>159.0±16.3</td> </tr> <tr> <td>Median</td> <td>1.78</td> <td>164.6</td> </tr> <tr> <td>LQ/HQ</td> <td>1.78/2.34</td> <td>134.0/174.1</td> </tr> </tbody> </table>		d [3-4] (Å)	angle [3-4-5] (°)	Mean	1.96±0.23	159.0±16.3	Median	1.78	164.6	LQ/HQ	1.78/2.34	134.0/174.1
	d [3-4] (Å)	angle [3-4-5] (°)													
Mean	1.96±0.23	159.0±16.3													
Median	1.78	164.6													
LQ/HQ	1.78/2.34	134.0/174.1													
8	<p>Posphine oxide / Posphone</p> <p>Hits/Fragments: 133/271</p>		<table border="1"> <thead> <tr> <th></th> <th>d [3-4] (Å)</th> <th>angle [3-4-5] (°)</th> </tr> </thead> <tbody> <tr> <td>Mean</td> <td>1.88±0.18</td> <td>162.3±15.9</td> </tr> <tr> <td>Median</td> <td>1.84</td> <td>166.4</td> </tr> <tr> <td>LQ/HQ</td> <td>1.73/2.0</td> <td>145.5/174.9</td> </tr> </tbody> </table>		d [3-4] (Å)	angle [3-4-5] (°)	Mean	1.88±0.18	162.3±15.9	Median	1.84	166.4	LQ/HQ	1.73/2.0	145.5/174.9
	d [3-4] (Å)	angle [3-4-5] (°)													
Mean	1.88±0.18	162.3±15.9													
Median	1.84	166.4													
LQ/HQ	1.73/2.0	145.5/174.9													
9	<p>Nitrile</p> <p>Hits/Fragments: 30/36</p>		<table border="1"> <thead> <tr> <th></th> <th>d [3-4] (Å)</th> <th>angle [3-4-5] (°)</th> </tr> </thead> <tbody> <tr> <td>Mean</td> <td>2.10±0.23</td> <td>160.4±17.4</td> </tr> <tr> <td>Median</td> <td>2.02</td> <td>167.0</td> </tr> <tr> <td>LQ/HQ</td> <td>1.85/2.60</td> <td>126.3/176.0</td> </tr> </tbody> </table>		d [3-4] (Å)	angle [3-4-5] (°)	Mean	2.10±0.23	160.4±17.4	Median	2.02	167.0	LQ/HQ	1.85/2.60	126.3/176.0
	d [3-4] (Å)	angle [3-4-5] (°)													
Mean	2.10±0.23	160.4±17.4													
Median	2.02	167.0													
LQ/HQ	1.85/2.60	126.3/176.0													



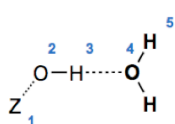
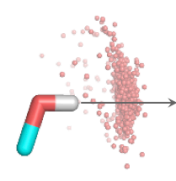
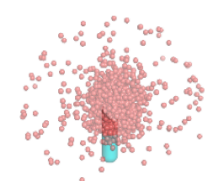
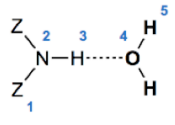
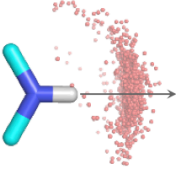
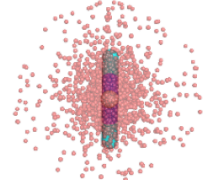
Water oxygen atoms interacting with hydroxyls and amino HBDs (Table 9 - entries 13, 14), are distributed along the two HEVs (OH and NH). Interestingly, in the case of water oxygen interacting with apolar hydrogens (Table 9 - entries 15, 16), the average median distance of 2.59 Å is lower than the sum of the van der Waals radii for hydrogen and oxygen of 2.72 Å. This value, together with the range between lower and higher quantile⁵ of the two queries (2.35 – 2.70 Å), shows that a weak interaction is established between the apolar hydrogen and the oxygen of the water and that the water oxygen position is not a consequence of artifacts or packing effects.¹⁴⁶ The localization of the oxygen atoms of water shows no preferred distribution. Based on atomic distances and lone pair deviations, it is deducible that water molecules are indeed interacting with these fragments through electrostatic interactions, and that their spatial orientation is mainly determined by the neighboring atoms.

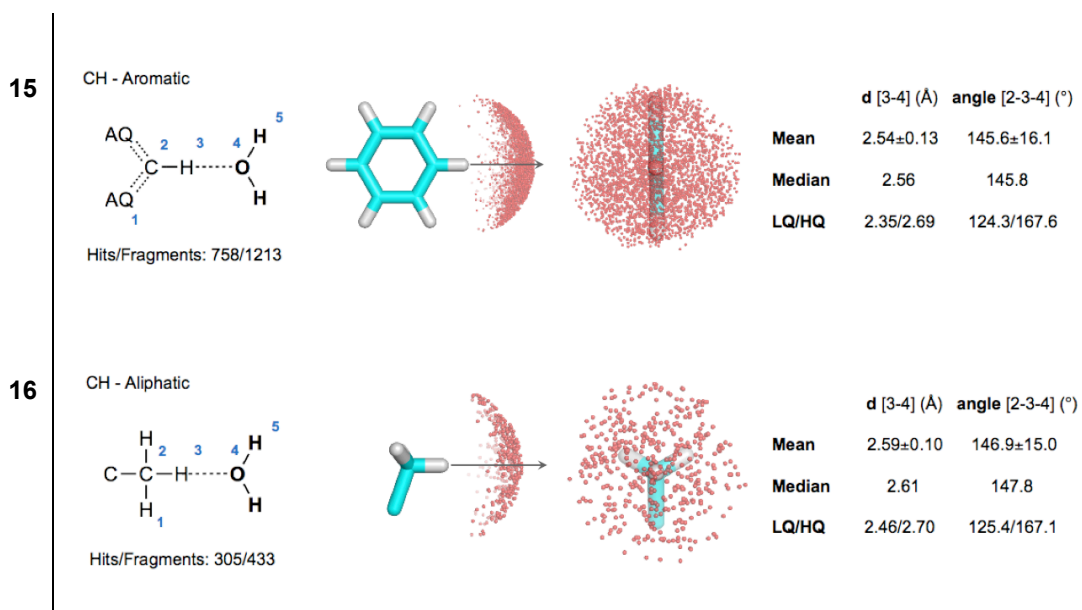
⁵ Quantiles: values marking equally sized consecutive subsets of an ordered sample of population (in our search analysis the quantile is set to 10).

One aim of the CSD search was to assess whether the small size and relatively high mobility of a water molecule has an impact on the geometry of hydrogen bonds. The results (Table S2 – Appendix 1 – Supplementary Information) are generally in agreement with values obtained from queries with a generic hydroxyl. However, a detailed analysis of all the interactions shows that water as an acceptor systematically forms shorter hydrogen bonds particularly when interacting with strong donors (OH, NH) when compared to the generic OH query, while water as hydrogen bond donor forms longer hydrogen bonds.

Thresholds and value ranges used in AcquaAlta are in agreement with the geometries obtained in the CSD search. In order to process protein structures with lower resolution and possibly less accurate geometries,^{181, 182} AcquaAlta softens the corresponding geometric criteria during an iterative scan.

Table 9: Summary of the crystallographic data search for functional groups acting as hydrogen bond donors (HBD). The organization is identical to Table 8 with the only exception of water oxygen atoms, which are displayed with red spheres.

<i>HBA: Functional groups acting as hydrogen bond acceptors</i>																
#	Query	Side view	Front view	Measurements												
13	OH - Hydroxyl  Hits/Fragments: 1266/1814			<table border="0"> <tr> <td></td> <td>d [3-4] (Å)</td> <td>angle [2-3-4] (°)</td> </tr> <tr> <td>Mean</td> <td>1.80±0.21</td> <td>162.70±15.0</td> </tr> <tr> <td>Median</td> <td>1.76</td> <td>166.7</td> </tr> <tr> <td>LQ/HQ</td> <td>1.6/1.99</td> <td>146.2/175.6</td> </tr> </table>		d [3-4] (Å)	angle [2-3-4] (°)	Mean	1.80±0.21	162.70±15.0	Median	1.76	166.7	LQ/HQ	1.6/1.99	146.2/175.6
	d [3-4] (Å)	angle [2-3-4] (°)														
Mean	1.80±0.21	162.70±15.0														
Median	1.76	166.7														
LQ/HQ	1.6/1.99	146.2/175.6														
14	NH - Amino  Hits/Fragments: 1138/1599			<table border="0"> <tr> <td></td> <td>d [3-4] (Å)</td> <td>angle [2-3-4] (°)</td> </tr> <tr> <td>Mean</td> <td>2.0±0.23</td> <td>156.7±16.6</td> </tr> <tr> <td>Median</td> <td>1.94</td> <td>160.8</td> </tr> <tr> <td>LQ/HQ</td> <td>1.75/2.34</td> <td>134.3/173.3</td> </tr> </table>		d [3-4] (Å)	angle [2-3-4] (°)	Mean	2.0±0.23	156.7±16.6	Median	1.94	160.8	LQ/HQ	1.75/2.34	134.3/173.3
	d [3-4] (Å)	angle [2-3-4] (°)														
Mean	2.0±0.23	156.7±16.6														
Median	1.94	160.8														
LQ/HQ	1.75/2.34	134.3/173.3														



1.2 Functional-group hydration propensity

Despite the availability of geometric criteria for describing hydrogen bonds (see Table 8 and 9), rules to calculate associated energies (enthalpy, entropy) are scarce.^{183, 184} In addition, the treatment of hydrogen-bond networks has only recently been addressed.^{3, 185} Up to date, the NIST database¹⁸⁶ list only a few experimentally determined parameters for strength of hydrogen bonds between water and small charged species like methylammonium and acetate.

Different hydration propensities of ligand functional groups and protein side-chains have been elucidated through an extensive statistical analysis performed on 392 high-resolution ligand–protein structures.¹³ A recent study¹⁸⁷ experimentally determined the relative basicity and therefore the strength of hydrogen-bond acceptors using Fourier transform infrared spectrometry (FTIR). When compared to NMR, UV and IR techniques, this method allows to identify and to rank HBAs of polyfunctional bases. Due to the extreme diversity of the structures deposited in the CSD, hydrogen-bond frequencies and geometries cannot be used to quantify their strengths for establishing quantitative structure-activity relationships. Nonetheless, the agreement between the abundance of certain hydrogen bonds in the CSD and the acceptor strength¹⁸⁷ can indicate the likelihood of hydrogen-bond formation.

In order to obtain relative interaction energies as well as geometric information, particularly valuable for interactions that are not well represented in the CSD, a series of *ab initio* calculations were performed on hydrated functional groups (Figure 25). Here, calculated hydrogen-bond interaction energies were used as a direct measure of a given functional group's propensity to hydration. A subsequent ranking based on those interaction energies inspired by a statistical analysis of 17 ligand atom types¹³ was used to fine tune a hydration-propensity scale for the AMBER* force-field atom types.¹⁴⁹

Oxygen atoms of carboxyls are most likely to be involved in hydrogen bond interaction. Their *ab initio* interaction energies compared to those of sp^3 oxygen atoms of ethers and especially esters are substantially higher (ΔE ranges from -14.7 to -5.2 kcal/mol vs. -5.0 to -3.3 kcal/mol). A similar trend for these two groups is observed when frequency of hits and number of fragments found in the CSD search are compared. Due to the low electronegativity of the phosphorus atom, the oxygen atom of a P=O group is also a strong acceptor ($\Delta E = -7.0$ kcal/mol), while the same atom in S=O or N=O groups forms comparably weaker hydrogen bonds (ΔE of -4.7 and -3.5 kcal/mol, respectively). A nitrogen sp^2 atom incorporated in pyridine rings forms a slightly stronger hydrogen bond than the sp hybridized nitrogen atom in acetonitrile. A sulfur atom, regardless of its hybridization state, was found to be a weak hydrogen bond acceptor, comparable to halogen atoms, where the fluorine atom was the most willing hydrogen bond acceptor. In the ideal case, aromatic carbon atoms can form a well defined σ - π interaction with a water molecule, contributing as much as -2.2 kcal/mol, which is roughly 40% of a strong hydrogen bond (e.g. $O-H\cdots O_{\text{acetone}} = -5.2$ kcal/mol).

When acting as a hydrogen-bond acceptor, water prefers to interact with charged nitrogen groups. Otherwise no substantial difference in interaction strength (ΔE) can be found between N-H and O-H groups, with both being frequent and potent hydrogen-bond donors. Thiols form only weak interactions with water oxygen ($\Delta E = -2.0$ kcal/mol), as a matter of fact, they are even weaker than the best interaction formed with hydrocarbons (e.g. for acetylene, $\Delta E = -2.45$ kcal/mol). The *ab initio* calculations further suggests that halogens form only weak interactions with water. The strongest halogen donor \cdots water interaction (Figure 25) was the one of bromobenzene featuring a ΔE of only -0.8 kcal/mol. Although, some of the listed interactions are quite rare (e.g. esters,⁴⁴ sulfones, and sulfonamides⁴⁹), they are included in Figure 25 to cover cases as found in some of the CSD structures.

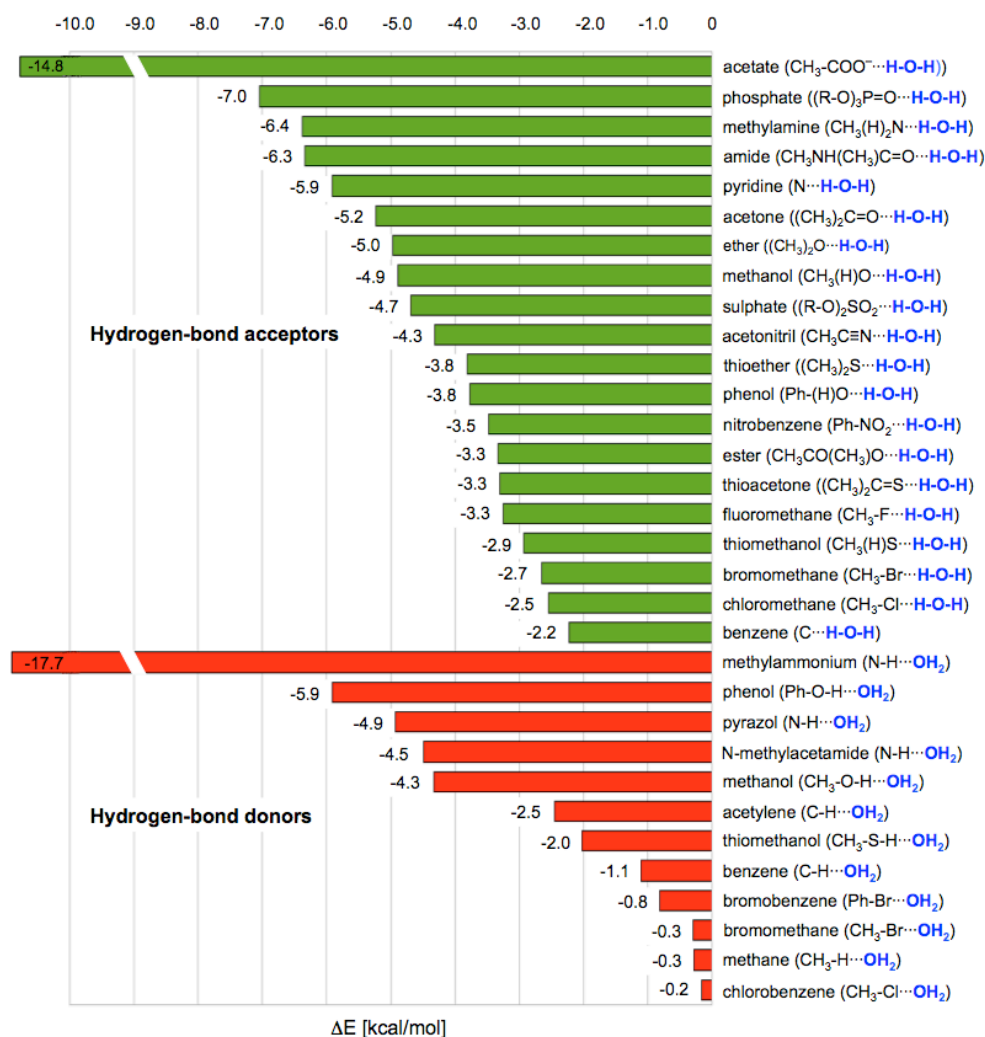


Figure 25: *Ab initio* calculated interaction energies for selected compounds interacting with water. Hydrogen-bond acceptors are depicted in green, hydrogen-bond donors in red.

1.3 Validation of AcquaAlta

The algorithm in AcquaAlta was validated on a dataset of suitable crystal structures retrieved from the PDB. The structures belong to different pharmacological and biological target families. The resolution of the selected 20 crystal structures (Table 6) ranges from 0.97 Å for the bovine beta-trypsin (PDB code: 2ayw) to 2.60 Å for the epidermal growth factor receptor (PDB code:

1m17). A subset of 15 structures (75%) showed resolutions lower or equal to 2.0 Å.

Calculated hydrogen atoms were added and oriented to protein and ligand structures prepared (as they are not resolved in the experimental structure) using the Yeti software. Then, the most probable protonation state at physiological pH (7.4) was calculated using Epik. In the validation, all experimental waters located within 3.25 Å from any ligand non-hydrogen atom (Figure 24a, indicated with arrows) were identified. To allow for an unbiased procedure, the experimental water coordinates were deleted (Figure 24b), and in the main step, were recalculated by AcquaAlta, oriented and minimized using the orient module in Yeti, based a directional algorithm. In the last step (Figure 24c), the positions of calculated waters were compared to the experimental position.

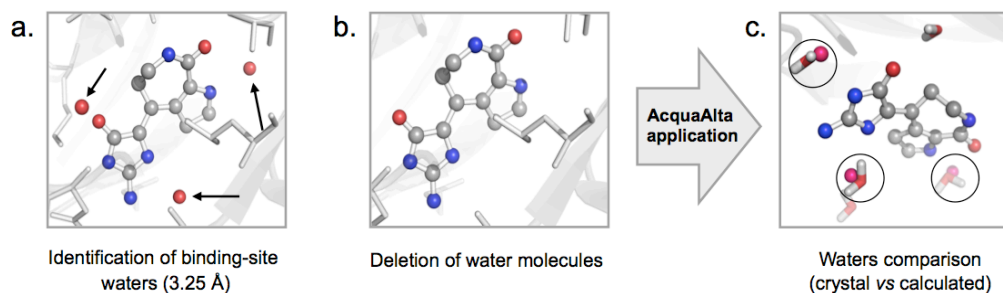


Figure 26: Algorithm validation. a) Identification of water molecules within 3.25Å from each heavy atom of the ligand. b) Deletion of the water molecules. Application of the AcquaAlta solvation algorithm (gray arrow). c) Comparison between experimental and calculated waters (oxygen atom only).

The match between experimental and calculated waters is summarized in Table 10. A filter (the criteria are listed in the Methods section) to discriminate between binding site waters and bridging waters was applied. The match refers to calculated waters compared with experimental waters for each structure. Different matches are provided when considering binding site waters (Table 10 – 2nd column) or bridging waters (Table 10 – 3rd column).

The comparison (Table 10 and Figure 27) shows a 76% match between experimental and calculated for binding-site waters, while the match for the identified bridging waters was 87.5%. Analysis of the subset of fifteen crystal structures with resolutions equal or lower to 2.0 Å provided a match of 76.2% for binding site waters and a better prediction of 87.8% for bridging waters, showing a small but not significant improvement of the results.

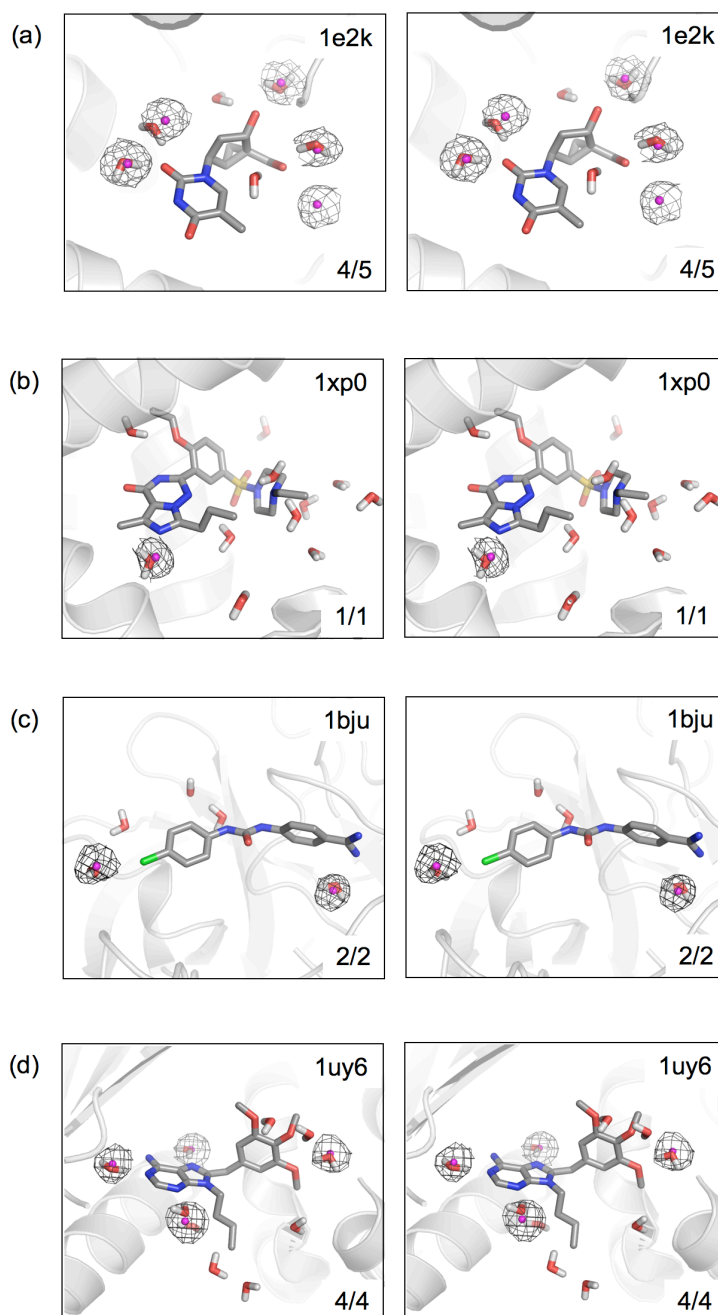
In six out of twenty structures (i.e. PDB codes: 3dfr, 1ywn, 1a81, 1hpx, 1u4d, 1kim), the match of the bridging waters is lower or equal to 66.7%. Two structures (3dfr, resolution 1.70 Å and 1a81, resolution 1.78 Å) contain rather high numbers of waters in the binding site (23 and 13, respectively). In both structures the high number of water molecules results from the presence of co-factors, which are also included in the solvation process, as well as by ligand localizations in solvent accessible areas. Despite a high resolution, only a low match (52.2% and 58.3% for binding site and bridging waters, respectively) is obtained for 3dfr, a fact that can probably be attributed to the localization of both ligand and co-factor on the enzyme surface where water networks can easily be formed. The poor result obtained for this structure can be explained by the fact that geometries related to water-networks are not implemented in our approach.

Selected results obtained in the validation phase (Table 10), are reported in Figure 27. In all the eight cases the match of the experimental binding site waters is higher or equal to the 60%. The calculated waters both match the experimental oxygen positions and are sterically fitting in the experimental water electron-densities.

Table 10: Results from the validation process. Listed are 20 PDB structures ordered by resolution (see Table 6). Examined are: (a) range of B-factor of all the water molecules present in each crystal structures, (b) number of binding site waters compared to the number of bridging waters, (c) match between experimental binding site waters and calculated waters, (d) weighted match between the filtered bridging waters and calculated waters. No direct correlation was found between water matching and B-factor range of the experimental waters.

#	PDB code	Binding site waters	Match	Bridging waters	Match
1	2ayw	8	6 (75%)	3	3 (100%)
2	3dfr	23	12 (52.2%)	12	7 (58.3%)
3	1e2k	5	4 (80%)	2	2 (100%)
4	1ywn	2	1 (50%)	1	1 (50%)
5	1a8i	13	10 (77.0%)	9	6 (66.7%)
6	1xp0	1	1 (100%)	1	1 (100%)
7	1bjv	2	2 (100%)	1	1 (100%)
8	1gd1	15	11 (73.3%)	8	6 (75%)
9	1uy6	4	4 (100%)	4	4 (100%)
10	1xgj	2	2 (100%)	2	2 (100%)
11	2b53	3	2 (66.7%)	1	1 (100%)
12	1o86	6	5 (83.4%)	4	4 (100%)
13	1h1d	4	3 (75%)	2	2 (100%)
14	1hpx	4	2 (50%)	3	2 (66.7%)
15	1ndw	6	3 (60%)	2	2 (100%)
16	1u4d	5	3 (60%)	3	2 (66.7%)

17	1f0r	2	1 (50%)	1	1 (100%)
18	1kim	3	2 (66.7%)	3	2 (66.7%)
19	1xkk	2	2 (100%)	1	1 (100%)
20	1m17	1	1 (100%)	1	1(100%)



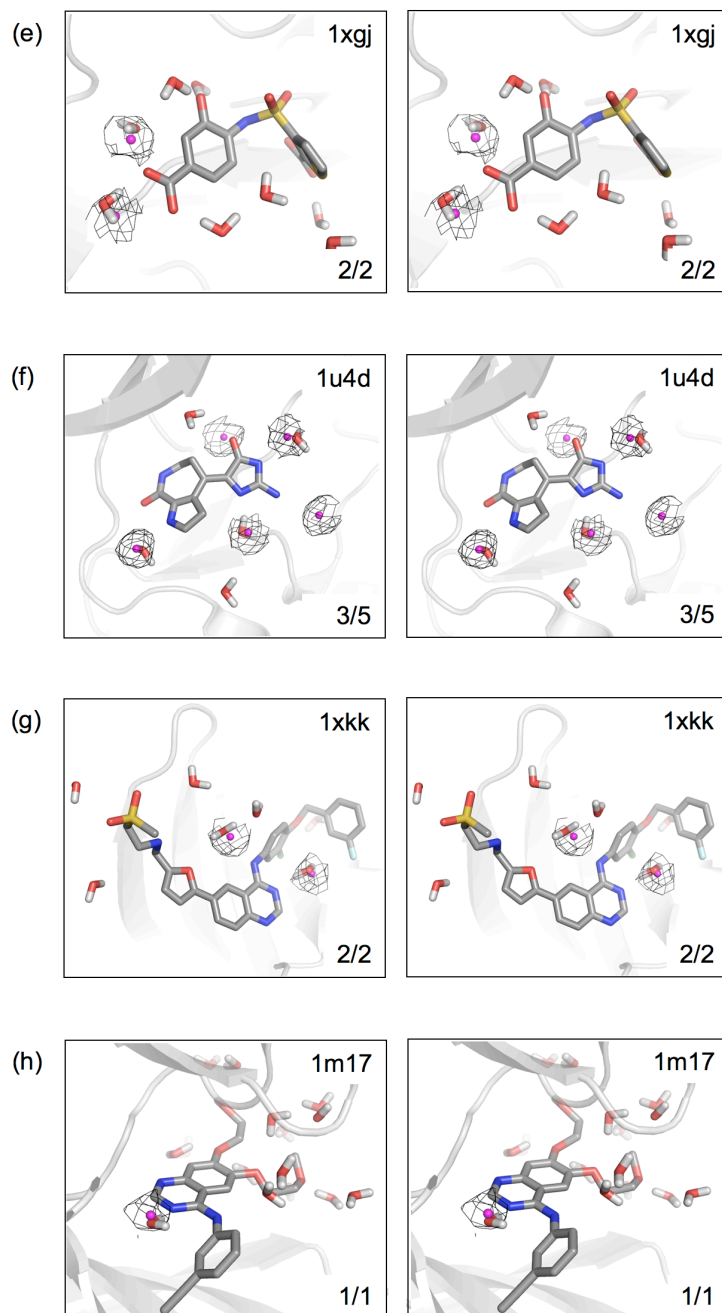


Figure 27: 3D representation (stereo view) of selected structures (PDB code in the top right corner) in the validation process. Ligand carbons are displayed in gray sticks, other ligand atoms are colored by atom type. Experimental waters are displayed as magenta spheres, water electron-density map as gray mesh and calculated waters in sticks representation. The match between experimental and calculated waters is given in the bottom right corner.

1.4 Algorithm application: Docking with Alignator

In order to challenge the solvation algorithm for the purpose of molecular docking, the oligopeptide binding protein (OppA) was chosen as target. Fourteen crystal structures of the OppA are available in the PDB with resolutions lower than 2.0 Å (Table 7). In these structures, OppA is complexed with different tripeptides of the type lysine-X-lysine (K-X-K), where X refers to any amino-acid residue.

Molecular docking was performed in two steps. First, the pre-alignment/docking tool Alignator¹⁵² was employed to align the tripeptides. The protocol aligns low-energy conformers of the docked molecule, obtained from conformational searching in aqueous solution, to a template molecule, based on pharmacophores matching. Solutions having unfavorable close contacts with the protein, which may request substantial induced fit for a proper accommodation, are removed.

For this purpose, the oligo-peptide KNK was selected as template. KNK contains the highest number of diverse pharmacophore centers. The structural similarity of the docked molecules and the template is high, however the scaffold KXK features a substantial number of rotatable bonds (>10). Moreover, for KXK motifs with X=D (aspartate) or X=E (glutamate), most of the low energy conformers have the two lysine side chains folded back to form an intramolecular salt bridge or hydrogen-bond stabilization. For this reason, the conformational search was set up with a rather large energy window (100 kcal/mol) and up to 12,000 conformers for each oligopeptide were retained and aligned by Alignator. Accepting a wider variety of conformers increases the probability of identifying extended conformer geometries similar to those bound in the binding pocket.

The oligopeptides aligned to KNK displayed a high match of pharmacophore centers and in the case of the “self-fit” (i.e. conformers of KNK aligned to itself) all the pharmacophores between the docked tripeptide and the template were matched. The *rmsd* of the matched pharmacophores (Table 11) is lower than 1.0 Å in all the cases except of KYK (Figure 28c) where the tyrosine side-chain adopts an alternative conformation compared to the crystal structure (χ_2 crystal = -61.43° , χ_2 docked = 58.36° , i.e. gauche⁻ vs gauche⁺) resulting in an apparent high *rmsd* value. The *rmsd* calculated on the heavy-atoms for aligned oligopeptides with the corresponding crystal conformation was always found to be lower than 1.0 Å, with the exception of KYK, KLK (Figure 28c,d), and KQK. High

rmsd for KYK and KQK is connected to the different side chain orientations of the X aminoacid. Alignments of tripeptides with long side chains (e.g. the two terminal lysine residues) are complex due to high number of possible combinations of their dihedral angles. Examples of aligned poses compared to the crystal structures, and experimental versus calculated water molecules, are shown in Figure 28.

After minimization, the water molecules generated with AcquaAlta were compared with the experimental waters for both the crystal and the Alignator-generated ligand poses, using the same distance criteria as in the validation (cf. above). Neither geometry nor energy filters were applied. The average match for the 14 crystal poses is 66%, which is by 10% lower than the match obtained for 20 complexes used in the validation. When the best *rmsd* pose from Alignator was used as input for the solvation algorithm, the resulting average match of water positions was 52.7%. A slight drop of accuracy can be expected since in the adopted alignment protocol, the conformers are aligned to the pharmacophore centers of a single oligo-peptide template (KNK), and as one can deduce from the *rmsd* scores, the matches are not perfect particularly for solvation sensitive hydrogen atoms in the terminal NH_3^+ group of the lysine side chain. Only steric clashes with the protein atoms are considered, but the position of specific protein hydrogen bond partners is not explicitly optimized.

Table 11: Results from automated docking: The 13 peptides aligned to the template KNK are listed along with the “self-fit” experiment (entry # 5). Included are: (a) number of the matched pharmacophore centers (PF) out of possible pharmacophores for each tripeptide, (b) the *rmsd* of the pharmacophores between template and aligned tripeptides, (c) the *rmsd* between the aligned tripeptide and the corresponding tripeptide from the crystal structure, (d) match of calculated to experimental waters when tripeptide from the X-ray structure and from the docking was used.

#	Peptide aligned to KNK	Matching Possible PF	<i>rmsd</i> PF	<i>rmsd</i> aligned to crystal	Match crystal	Match aligned pose
1	KAK	26/28	0.486	0.623	4/7 (57.1%)	4/7 (57.1%)
2	KEK	27/31	0.768	0.634	5/9 (55.6%)	6/9 (66.7%)
3	KWK	27/31	0.923	0.916	5/6 (83.3%)	3/6 (50%)
4	KDK	28/30	0.805	0.619	5/10 (50%)	4/10 (40%)
5	KNK	32/32	0.550	0.495	5/7 (71.4%)	3/7 (42.9%)

6	KMK	26/31	0.811	0.513	3/7 (42.9%)	3/7 (42.9%)
7	KHK	28/32	0.975	0.933	5/7 (71.4%)	2/7 (28.6%)
8	KPK	24/25	0.597	0.831	5/6 (83.3%)	5/6 (83.3%)
9	KSK	28/29	0.721	0.660	7/9 (77.8%)	6/9 (66.7%)
10	KYK	27/30	1.068	2.885	5/7 (71.4%)	3/7 (42.9%)
11	KQK	30/33	1.271	1.280	6/10 (60%)	5/10 (50%)
12	KLK	27/30	0.839	1.230	5/6 (83.3%)	4/6 (66.7%)
13	KRK	27/37	0.533	0.855	5/6 (83.3%)	2/6 (33.3%)
14	KVK	25/30	0.768	0.540	2/6 (33.3%)	4/6 (66.7%)

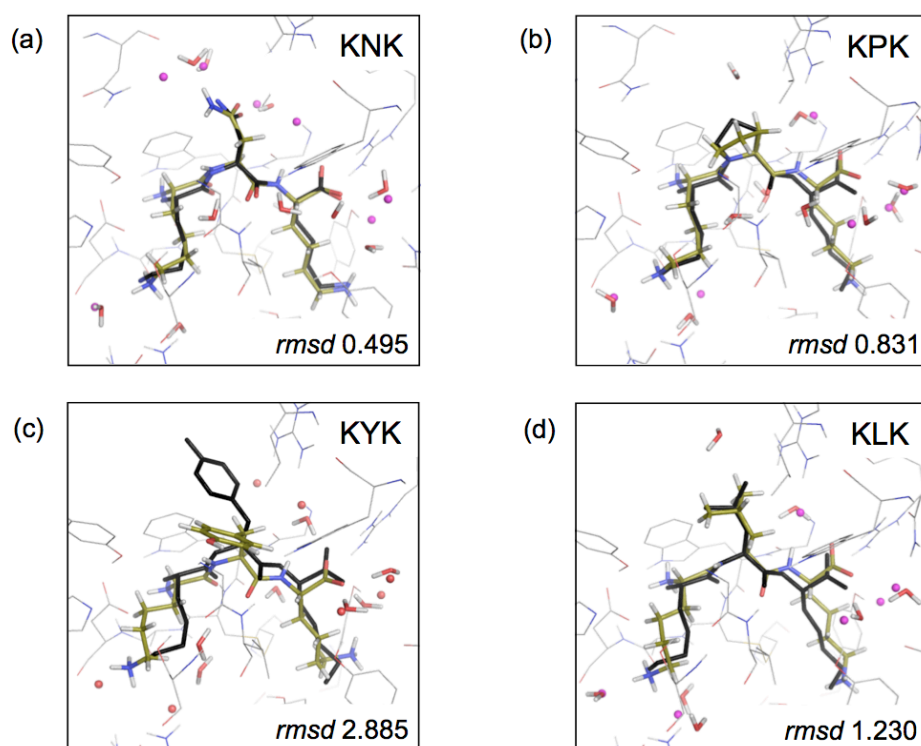


Figure 28: 3D representation (stereo view) of selected PDB structures (tripeptide sequence, upper right) during the automated docking process. Crystal ligands are displayed in black sticks without oxygen atoms. Experimental waters are displayed as magenta spheres and calculated waters, based on the docked result, in sticks representation. In lower right the heavy-atoms *rmsd* of the oligopeptide (conformations between the crystal and the docked result).

1.5 Comparison with the solvation module of Yeti

The solvation module of Yeti¹⁴⁴ is conceptually similar to AcquaAlta. The most important shared features are the accounting of the directionality of the hydrogen bonds as well as the use of vectors to display lone-pair positions as well as the vectors as extensions from the hydrogen bond donors. AcquaAlta is based on an updated CSD search both on functional groups acting as hydrogen bond acceptors and donors, and on the interaction of waters with halogen atoms, as well as with hydrophobic ligands moieties.

In order to test if the performances of AcquaAlta were comparable to the ones of the Yeti module, the validation procedure on the 20 X-ray structures (see Results 1.2) was performed using the module with “Standard” and “Tight” settings, respectively.

The comparison between the two performances displayed better results in term of water matching from AcquaAlta compared to the Yeti module both in the case of the match of binding-site waters and bridging waters (Table 12, 13).

“Standard” and “Tight” settings in the module refer to the parameters used in the cavity scan. For “Standard” the vdW value is 0.891 (threshold value below which the Leonard-Jones 6-12 potential is repulsive) and the hydrogen bond value is 0.913 (threshold value below which the 10-12 potential¹⁸⁸ is repulsive); in the case of “Tight” these values are scaled down multiplying them by a factor of 0.8. During the solvation with the solvation module of Yeti single H-bonded waters were included in the solvation procedure.

Table 12: Performance comparison in water matching for AcquaAlta and the AUTO-SOL module of Yeti with “Standard” settings. The 20 X-ray structures are identified by PDB codes, BSW stands for binding-site waters, BW stands for bridging waters. When AcquaAlta performance was superior to AUTO-SOL a bold plus sign is used, when inferior a bold minus sign (–) was used, when equal a bold equal sign is used (=).

	PDB	AcquaAlta BSW Match	BW Match	AUTO-SOL BSW Match	Standard BW Match
1	2ayw	6/8 (75%) +	3/3 (100%) +	0/8(0%)	0/3(0%)
2	3dfr	12/23 (52.2%) +	7/12 (58.3%) +	4/23(17.4%)	3/12(25%)
3	1e2k	4/5 (80%) +	2/2 (100%) +	3/5(60%)	1/2(50%)
4	1ywn	1/2 (50%) +	1/2 (50%) +	0/2(0%)	0/1(0%)
5	1a8i	10/13 (77.0%) +	6/9 (66.7%) +	3/13(23.1%)	2/9(22.2%)
6	1xp0	1/1 (100%) +	1/1 (100%) =	1/1(100%)	1/1(100%)
7	1bjv	2/2 (100%) +	1/1 (100%) +	0/2(0%)	0/1(0%)
8	1gd1	11/15 (73.3%) +	6/8 (75%) +	4/15(26.7%)	3/8(37.5%)

9	1uy6	4/4 (100%) +	4/4 (100%) +	1/4(25%)	1/4(25%)
10	1xgj	2/2 (100%) +	2/2 (100%) +	1/2(50%)	1/2(50%)
11	2b53	2/3 (66.7%) +	1/1 (100%) +	0/3(0%)	0/1(0%)
12	1o86	5/6 (83.4%) +	4/4 (100%) +	0/6(0%)	0/4(0%)
13	1h1d	3/4 (75%) +	2/2 (100%) +	0/4(0%)	0/2(0%)
14	1hpx	2/4 (50%) +	2/3 (66.7%) +	0/4(0%)	0/3(0%)
15	1ndw	3/5 (60%) +	2/2 (100%) +	1/6(16.7%)	1/2(50%)
16	1u4d	3/5 (60%) +	2/3 (66.7%) +	1/5(20%)	1/3(33.3%)
17	1f0r	1/2 (50%) +	1/1 (100%) +	0/2(0%)	0/1(0%)
18	1kim	2/3 (66.7%) +	2/3 (66.7%) +	1/3(33.3%)	1/3(33.3%)
19	1xkk	2/2 (100%) +	1/1 (100%) +	0/2(0%)	0/1(0%)
20	1m17	1/1 (100%) +	1/1(100%) +	0/1(0%)	0/1(0%)

Table 13: Performance comparison in water matching for AcquaAlta and the AUTO-SOL module of Yeti with “Tight” settings. The 20 X-ray structures are identified by PDB codes, BSW stands for binding-site waters, BW stands for bridging waters. When AcquaAlta performance was superior to AUTO-SOL a bold plus sign is used (+), when inferior a bold minus sign (–) was used, when equal a bold equal sign is used (=).

	PDB	AcquaAlta BSW Match	BW Match	AUTO-SOL BSW Match	Tight BW Match
1	2ayw	6/8 (75%) +	3/3 (100%) +	3/8(37.5%)	2/3(66.7%)
2	3dfr	12/23 (52.2%) =	7/12 (58.3%) +	12/23(52.2%)	5/12(41.7%)
3	1e2k	4/5 (80%) +	2/2 (100%) +	2/5(40%)	1/2(50%)
4	1ywn	1/2 (50%) –	1/2 (50%) –	2/2(100%)	1/1(100%)
5	1a8i	10/13 (77.0%) =	6/9 (66.7%) =	10/13(77.0%)	6/9(66.7%)
6	1xp0	1/1 (100%) =	1/1 (100%) =	1/1(100%)	1/1(100%)
7	1bjv	2/2 (100%) +	1/1 (100%) +	0/2(0%)	0/1(0%)
8	1gd1	11/15 (73.3%) +	6/8 (75%) +	8/15(53.3%)	4/8(50%)
9	1uy6	4/4 (100%) +	4/4 (100%) +	3/4(75%)	3/4(75%)
10	1xgj	2/2 (100%) +	2/2 (100%) +	0/2(0%)	0/2(0%)
11	2b53	2/3 (66.7%) +	1/1 (100%) =	3/3(100%)	1/1(100%)
12	1o86	5/6 (83.4%) +	4/4 (100%) +	3/6(50%)	2/4(50%)
13	1h1d	3/4 (75%) +	2/2 (100%) +	1/4(25%)	1/2(50%)
14	1hpx	2/4 (50%) =	2/3 (66.7%) =	2/4(50%)	2/3(66.7%)
15	1ndw	3/5 (60%) +	2/2 (100%) =	2/6(33.3%)	2/2(100%)
16	1u4d	3/5 (60%) +	2/3 (66.7%) +	2/5(40%)	1/3(33.3%)
17	1f0r	1/2 (50%) +	1/1 (100%) +	0/2(0%)	0/1(0%)
18	1kim	2/3 (66.7%) –	2/3 (66.7%) –	3/3(100%)	3/3(100%)
19	1xkk	2/2 (100%) +	1/1 (100%) +	1/2(50%)	0/1(0%)
20	1m17	1/1 (100%) +	1/1(100%) +	0/1(0%)	0/1(0%)

2 MeTOX: Metabolism and TOXicology prediction

The first part of the study consisted on correlating the atomic thermal motion from the X-Ray crystal structure of CYP2D6 and CYP2C9 to the atomic movements observed during MD simulations. Based on the found correlation ligand pre-alignment and automated docking were used to establish a mQSAR study. To challenge the models, the consistency with sites of metabolism, the sensitivity to biological data (Y-scrambling), together with the prediction of binding and non-binding compounds were checked. Finally, the validated models for CYP2D6 and CYP2C9 were added to in the VirtualToxLab.

2.1 Correlation of experimental and calculated atomic motion

Plasticity, *i.e.* the ability of the active site of a target protein to accommodate and adapt to numerous different substrates,⁸⁸ and protein flexibility are properties shared by several members of the CYP family.^{88, 91} These aspects are often utilized to explain their high substrate-binding promiscuity along with the high regioselectivity.

CYP2D6 and CYP2C9 share more than 40% of their primary sequence and are similar in folding and organization of their secondary structural elements.¹⁵³ Still, only 18 out of 56 compounds binding to CYP2D6 also bind to CYP2C9 with affinities < 200 μM (Table 1, Appendix 2 – Supporting Information).

To be consistent with the hypothesis of different states of the CYP2C9 active site¹⁸⁹ and to explain the regioselectivity of binding, a realistic simulation must allow for induced-fit adaptation.⁸⁸ Moreover, the active-site volume calculated for the CYP2D6 structure is too small to allocate some of the known substrates.¹⁵³ In order to assess the appropriateness of our protein-relaxation protocol, the average displacement (\bar{u}) derived by the thermal displacement factors of the C α carbons (as obtained from the X-ray crystal structure determination) was compared to the average *rms* from the MD simulations (Figure 29 and 30).

In CYP2D6 (Figure 29), the C α carbons of the amino acids in the core of the active site display a rather high stability both in terms of \bar{u} and average *rms*. Besides the low \bar{u} and average *rms* of key residues lining the active site (*e.g.*

Phe120, Val370), Glu216, Asp301 and Phe483 are located in a more flexible area (indicated by an asterisk in Figure 3). A larger mobility is observed in the loops B–C, and F–G located at the protein surface compared to loops G–H, H–I and β sheet β 2–1.

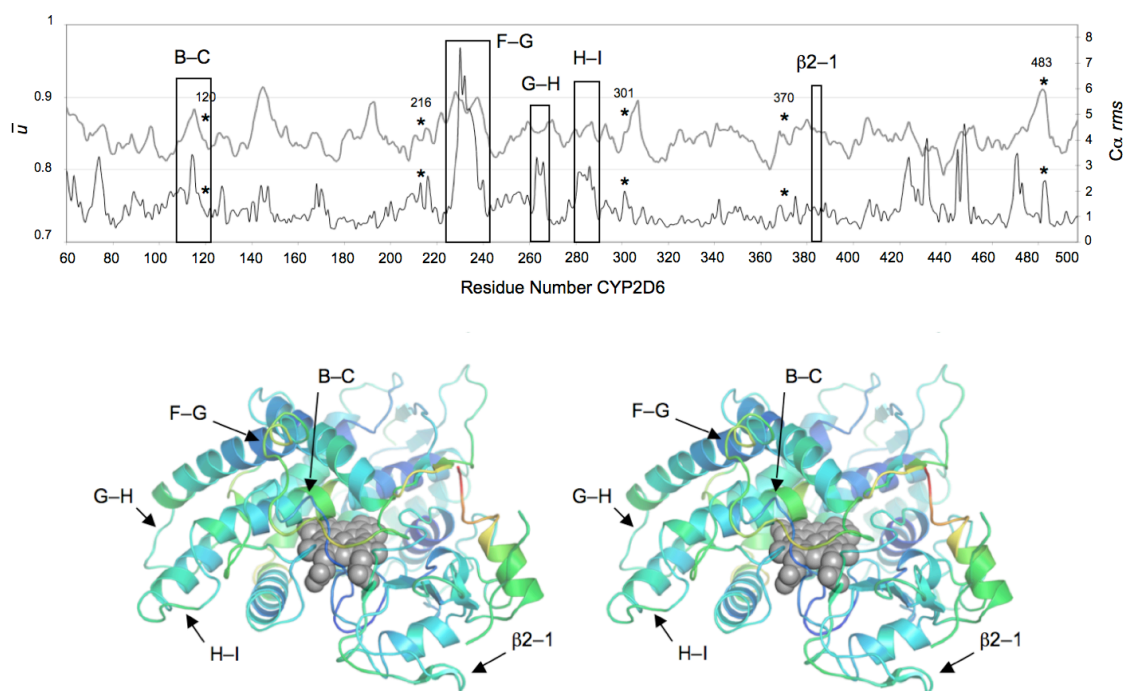


Figure 29: CYP2D6 – Average displacement \bar{u} (X-Ray; left y-axis, light gray, upper curve), and calculated *rms* (MD; right y-axis, dark gray, lower curve) for CYP2D6 after alignment and superposition of 1,000 frames recorded along an MD simulation of 1.2 ns for each $C\alpha$ carbon of each residue (x-axis). Different ranges are used between \bar{u} and $C\alpha$ *rms* to highlight the similar trends. 3D representation (stereo view) of CYP2D6 with $C\alpha$ carbons colored by the B factor and the heme portion displayed in space-filling mode. Loops at the protein surface, and the positions of residues important for the binding are highlighted (*).

In CYP2C9 (Figure 30), high mobility is displayed by several loops (F–G, G–H, H–I, J–J¹), and a β sheet (β 2–1) at the protein surface. As already emphasized,¹⁸⁹ the active site is rather rigid and the majority of its residues (e.g. Asp293, Thr301) present low values of both \bar{u} and average *rms*. In comparison, in terms of \bar{u} , Glu104, Arg108, and Phe120 are more flexible. Finally, the mobility information for Glu104 and Arg108 largely differs between \bar{u} and calculated *rms*.

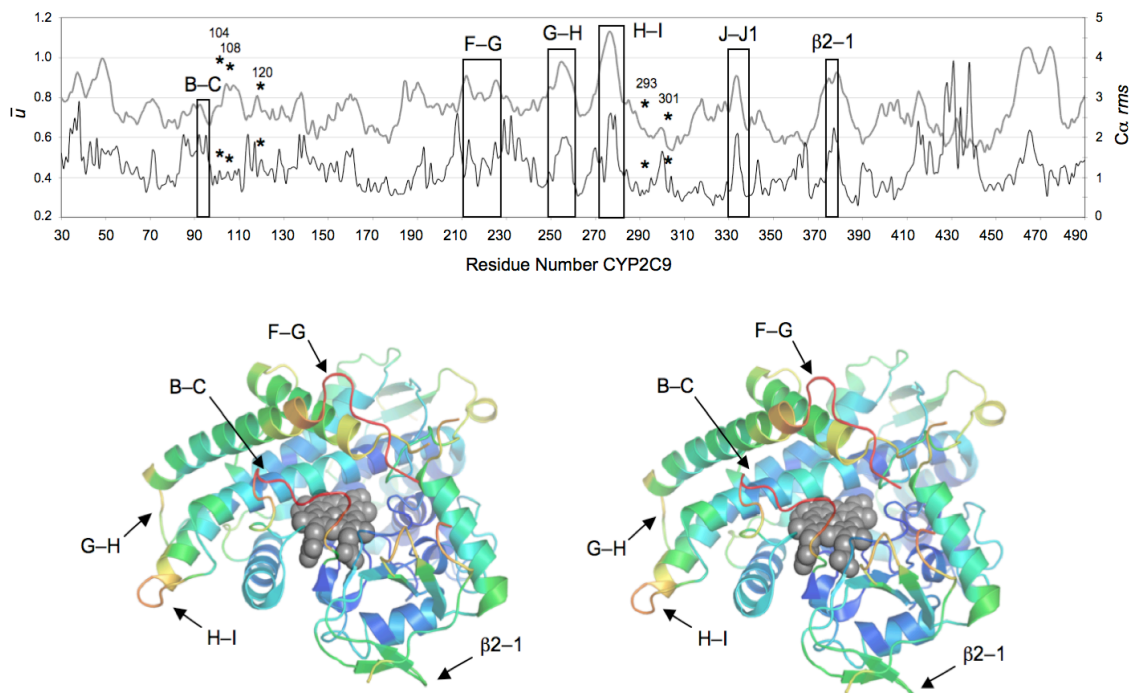


Figure 30: CYP2C9 – Average displacement \bar{u} (X-Ray; left y-axis, light gray, upper curve), and calculated rms (MD; right y-axis, dark gray, lower curve) for CYP2C9 after alignment and superposition of 1,000 frames recorded along an MD simulation of 1.2 ns for each $C\alpha$ carbon of each residue (x-axis). Different ranges are used between \bar{u} and $C\alpha rms$ to highlight the similar trends. 3D representation (stereo view) of CYP2C9 with $C\alpha$ carbons colored by the B factor and the heme portion displayed in space-filling mode. Loops at the protein surface, and the positions of residues important for the binding are highlighted (*).

This discrepancy between MD simulations and X-ray structural information could arise from an incorrect handling of the system by the force field. Furthermore, the length of the simulation (1.2 ns), simulation conditions and water model used to mimic the X-ray determination conditions, could also contribute to the observed discrepancy. Of course, in the case of the mutated residues that were reset to the wild type (e.g. D230L in CYP2D6, Figure 29), some inconsistency is to be expected. The aforementioned comparison only offers qualitative information on enzyme motion due to numerous limiting factors as atoms selection (*i.e.* only $C\alpha$ carbons), atom position (*i.e.* atoms belonging to loops and at the enzyme surface presents higher fluctuations), and difficulties in comparing absolute values for single atoms out of the two experiments (X-ray and MD simulation). Nonetheless, the numerous agreements allow a general interpretation of enzyme motion in a true biological system.

2.2 Prealignment and automated, flexible docking

The correlation of \bar{u} and average *rms* (Figure 29 and 30) further underlines the role of protein flexibility and the requirement of a flexible docking protocol simulating the binding to CYP2D6 and CYP2C9. Previous studies^{84, 190, 191} provided evidence that, despite the high promiscuity in ligand binding of CYPs, common pharmacophore features can be identified for the different isoforms. In order to include this binding information in our approach, we have adopted a pharmacophore-based pre-alignment, using methdilazine as template for CYP2D6 (Figure 31) and trifluoperazine, medrysone, piroxicam, and F02 for CYP2C9 (Figure 32). Since we identified only one chemical class (*i.e.* tricyclics) for CYP2D6, a unique template was used for the dataset. For CYP2C9 the compounds displaying the strongest binding affinity among the three chemical classes identified (*i.e.* tricyclics, steroids, sulfonamides) were used as templates. Piroxicam and F02 fit to the sulfonamide-like ensemble but were used separately, because they display distinct pharmacophore features.

Ligand pre-alignment was followed by a combinatorial scan of amino acid side-chains in the active site to identify possible hydrogen bond partners and to rearrange the active site consequently. This active site configuration represented the starting point for our flexible-docking protocol as implemented in Cheetah.^{160, 167} In this approach, the protein backbone is kept rigid but amino-acid side-chains are allowed to move. The docking output based on the pre-aligned conformations was energetically scored towards docking poses obtained through an extended Monte Carlo search. For 75% of the CYP2D6 and 61% of the CYP2C9 ligands, at least one structure obtained from the pre-alignment by Alignator¹⁵¹ is retained in the ensemble of the best-scored poses, which justifies usage of both pre-alignment algorithm and the Monte Carlo searching. A flowchart describing the pre-alignment and docking protocol is provided in Figure 21.

Methdilazine docked to CYP2D6 (Figure 31) is accommodated in a hydrophobic pocket lined by Leu213, Leu302, Phe120 and Ala305. Its tricyclic moiety is located in the close vicinity of the heme, possibly interacting through π - π interactions. The positively charged nitrogen atom which forms a strong electrostatic interaction with Glu216 (*cf.* Figure 31), together with the aromatic ring located close to the heme, correctly resemble the pharmacophore hypothesis for CYP2D6.⁸⁴ Accessibility of the aromatic carbon atoms of the ligand (highlighted in Figure 31) by the iron co-factor of the heme agrees with

the experimental data,¹⁹² which identify these atoms of tricyclic compounds as privileged sites of metabolism.

Previous site-directed mutagenesis experiments underline the importance of residues Phe120, Glu216, Asp301 for binding.^{162, 163} All of them are involved in the binding mode, especially Glu216 which interacts with ligands containing a positively charged nitrogen at physiological pH (47 out of 56 ligands). In principle, the negatively charged residue Asp301 could also face the active site and interact with the positive nitrogen from diverse ligands. However, a structural role in protein folding and heme incorporation¹⁹³ is attributed to this residue. Finally, Phe120 was shown to influence regioselectivity of substrate oxidation (e.g. the mutant F120L increases K_m for dextromethorphan by a factor of 30).¹⁶³

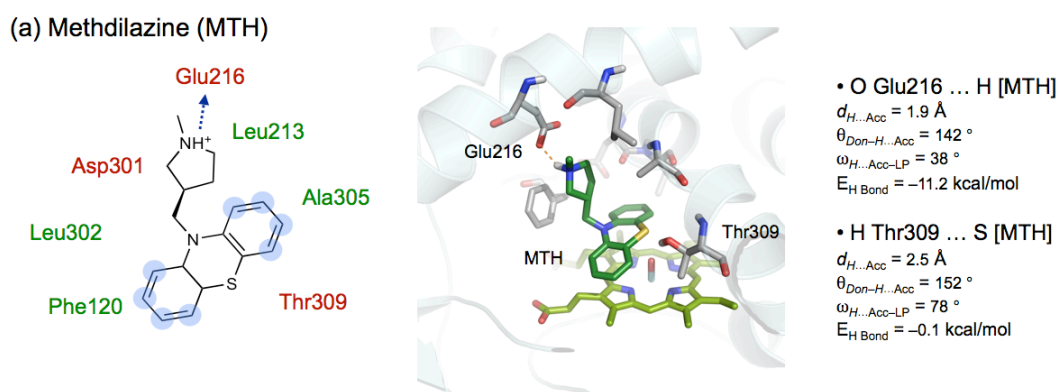


Figure 31: CYP2D6 – Details of binding for the template methdilazine used for the pre-alignment. In the left column, a 2D representation of the active site is shown: polar amino-acids are depicted in red; hydrophobic amino-acids in green, atoms of the ligand exposed and not interacting with any residue of the enzyme are blurred in blue. In the middle column, a 3D representation of the binding mode is given. Residues of the binding site are shown as gray, methdilazine as dark green and the heme as light-yellow sticks. In the right column, details of the main interactions in terms of distance(d), angle(θ), lone-pair deviation(ω) and energy contribution($E_{H \text{ Bond}}$; in kcal/mol) between methdilazine and CYP2D6 are listed.

The four templates (trifluoperazine, medryson, piroxicam and F02 - Figure 32) bind to a similar area of the active site of CYP2C9. The hydrophobic residues (Val113, Leu208, Ala297) are lining the channel leading to the heme and contribute hydrophobic interactions in the different binding modes of the four templates. In the trifluoperazine complex with CYP2C9 (Figure 32a), the electrostatic interaction between the positively charged nitrogen of the ligand

and Asp293 acts as anchor orienting the aromatic rings of the ligand orthogonal to the side-chain of Phe476.

The binding mode proposed for medrysone (Figure 32b) locates the rigid chemical scaffold farther away from the heme system. Two hydrogen bonds are established between ligand and protein: (1) between the carbonyl of ring A and the backbone NH of Leu102 and (2) between the oxygen atom of the hydroxyl group of the ligand and the side-chain of Arg108.

Despite the fact that piroxicam and F02 (Figure 32c, 32d) share a sulfonamide group, they have been selected for additional functional groups: amide and pyridinium moiety in piroxicam, a sulphur atom and a second sulfonamide in F02. Piroxicam establishes a hydrogen bond between the sulfonamide and the side-chain of Arg108, and an electrostatic interaction between the pyridinium moiety and the side-chain carboxylate of Asp293. One sulfonamide of F02 similarly interacts with the side-chain of Arg108 allowing the aromatic ring, located at the edge of ligand side-chain, to establish a close contact with the iron atom of the heme.

Similarly to the finding for CYP2D6, the privileged site of metabolism by CYP2C9 of the ligands trifluoperazine,¹⁹⁴ medrysone,¹⁹⁵ and piroxicam¹⁶⁴ are considered to be exposed to and accessible for the iron in the proposed binding modes (Figure 32 a–d). No data have been found for F02. Nonetheless, due to the chemical similarity with piroxicam, we can hypothesize that the site of metabolism is located on the aromatic ring docked in a position accessible to the heme.

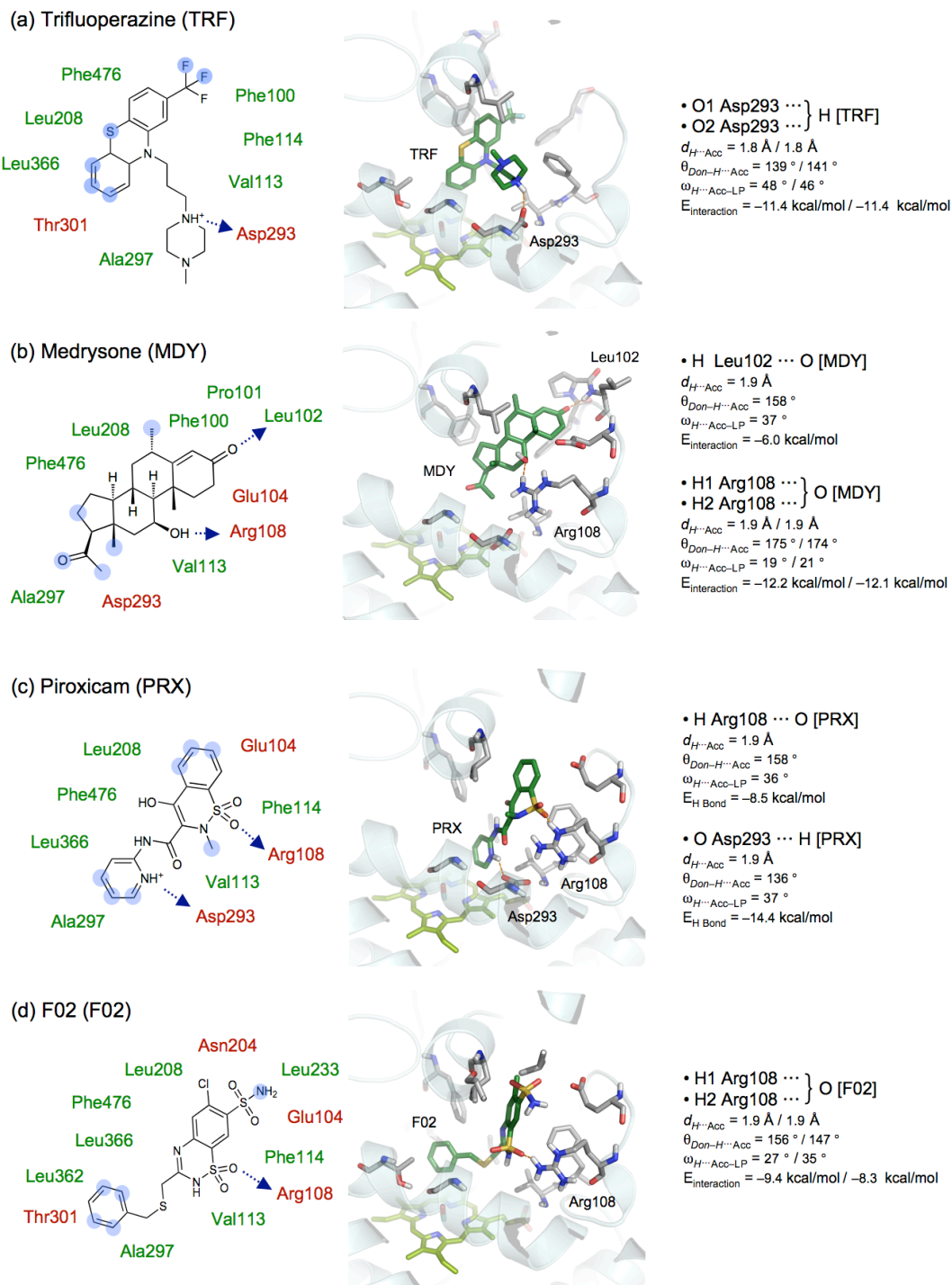


Figure 32: CYP2D9 – Binding details of the pre-alignment templates trifluoperazine (a), medryson (b), piroxicam (c) and F02 (d). For organization, color codes, representations and abbreviations cf. Figure 29.

2.3 Multidimensional QSAR (mQSAR)

Up to four poses within an energy window of 10 kcal/mol were selected for each of the 56 ligands of CYP2D6 and 85 ligands of CYP2C9, including different stereoisomers and protonation states for a total of 216 poses for CYP2D6 and 351 poses for CYP2C9. The 4D ensemble of the docked poses was used as input for Quasar.^{106, 107}

The model family of CYP2D6 (Figure 33a and Table 14) converged at cross-validated r^2 of 0.811 for the 46 training compounds leaving one third of the set out, and yielded a predictive r^2 of 0.711 for the 10 test ligands. The average deviation (*rms*) between experimental and calculated affinities is of a factor of 1.2 for the training and 2.4 for the test set. The maximal deviation in the prediction of binding affinities for a compound is of a factor of 5.0 for training set and 8.5 for the test set, respectively.

The model family of CYP2C9 (Figure 33b and Table 14) converged at a cross-validated r^2 of 0.687 for the 68 training compounds leaving one fifth of the set out. The predictive r^2 of 0.423 for the 17 test ligands is 0.423. The p^2 value is only moderate due to the distribution of the affinities for CYP2C9 together with the rather narrow range of affinities (3.25 orders of magnitude). The calculation of the regression coefficient (q^2) is mainly influenced by the affinities distribution and from the corresponding deviation from the average experimental affinity value as in equation 2. The predictive r^2 (test set) is also influenced by the distribution and range of affinities of the test set compounds (Equation 4). Thus, the relatively modest q^2 value is not a consequence of poor predictivity but of the limited data available.

$$q^2 = 1.0 - \frac{\sum(\Delta G_{pred} - \Delta G_{exp})^2}{\sum(\Delta G_{exp} - \Delta G_{exp,average})^2} \quad (\text{Eqn. 4})$$

This unavoidable inhomogeneity of distribution is visible in Figure 33b where the majority of ligands bind in the low-affinity range. This observation further correlates with the promiscuous character of these isoforms of enzymes. Despite the mentioned values, the worst predicted compound deviates from the experimental value by only a factor of 4.4 for the training set and of a factor of 4.8 for the test set. The average *rms* consistently reflects a low deviation between experimental and predicted binding affinity both for training and test set diverging by a factor of 1.2 and 1.7, respectively.

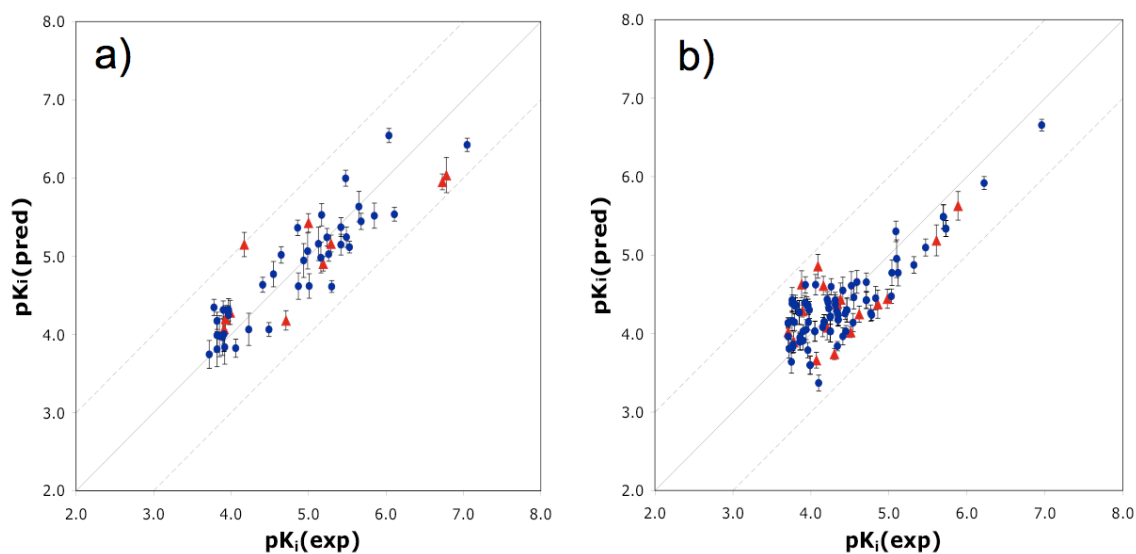


Figure 33: Comparison of experimental and predicted binding affinities of the training set (open circles) and test set (filled triangles) for CYP2D6 (a) and CYP2C9 (b). Dashed lines are drawn at a factor 10 from the experimental value.

Table 14: Summary of the Quasar results. *rms* training and *rms* test stand for average *rms* deviation from the experimental value for training and test rest respectively. Max. *rms* training and max. *rms* test stand for the maximum deviation from the experimental value noticed for training and test set respectively. *rms* and maximal values are expressed in factors of K_i .

	r^2	q^2	<i>rms</i> training	max. training	p^2	<i>rms</i> test	max. test
CYP2D6	0.815	0.811	1.2	5.0	0.711	2.4	8.5
CYP2C9	0.695	0.687	1.2	4.4	0.423	1.7	4.8

q^2 = cross-validated r^2

p^2 = predictive r^2

2.4 Consistency with the sites of metabolism

Attempting to evaluate the reliability of the docking protocol and the subsequent quantification by mQSAR, the best-ranked poses for some of the ligands in the dataset were visually inspected. The aim was to check the accessibility of the experimentally known sites of metabolism, for the iron atom of the heme during a hypothetical metabolic hydroxylation process.

In Figure 34a–c, the vicinity of the sites of metabolism to the heme for three compounds binding to CYP2D6 are shown. CYP2D6 catalyzes the transformation of nortriptyline (Figure 34a) to its metabolite 10-hydroxynortriptyline through a hydroxylation process.¹⁹⁶ Within the docking selection (see selection criteria above) the carbon atom in position ten is accessible to the heme; similarly the sites of metabolism for other two tricyclic antidepressants (clomipramine and imipramine), propranolol and ondansetron¹⁹¹ are matched and accessible in the docking poses obtained from this protocol. O-demethylation is another metabolic reaction processed by CYP2D6. Galantamine (Figure 34b) undergoes this process and its methoxy group is metabolized by oxidative dealkylation into the corresponding hydroxyl group.¹⁹⁷ In the docking pose the vicinity of this functional group to the heme is consistently found. Clomipramine (Figure 34c), similarly to nortriptyline, can be hydroxylated in position 8,¹⁹⁸ which is accessible to heme in the docking pose obtained.

The sites of metabolism of tolbutamide, dapsone, losartan, chlorpropamide, diclofenac, phenytoin¹⁹⁹ are linearly accessible for the heme of CYP2C9. However, the distances from the oxygen atom of the iron-oxygen system is rather high (e.g. dapsone 4.2 Å, diclofenac 5.2 Å, phenytoin 7.6 Å). One reason could be the large volume of the active site. In addition, in the X-ray structure of CYP2C9 in complex with warfarin, the ligand is located in a distal part of the active site 10 Å away from the iron of the heme.⁹⁰ Williams et al. suggested that this secondary pocket could represent a primary recognition site or a binding pocket for competitive inhibition of the enzyme. In Figure 34d–f are shown examples of the docking poses and their relative position to the heme for dapsone, diclofenac and phenytoin.

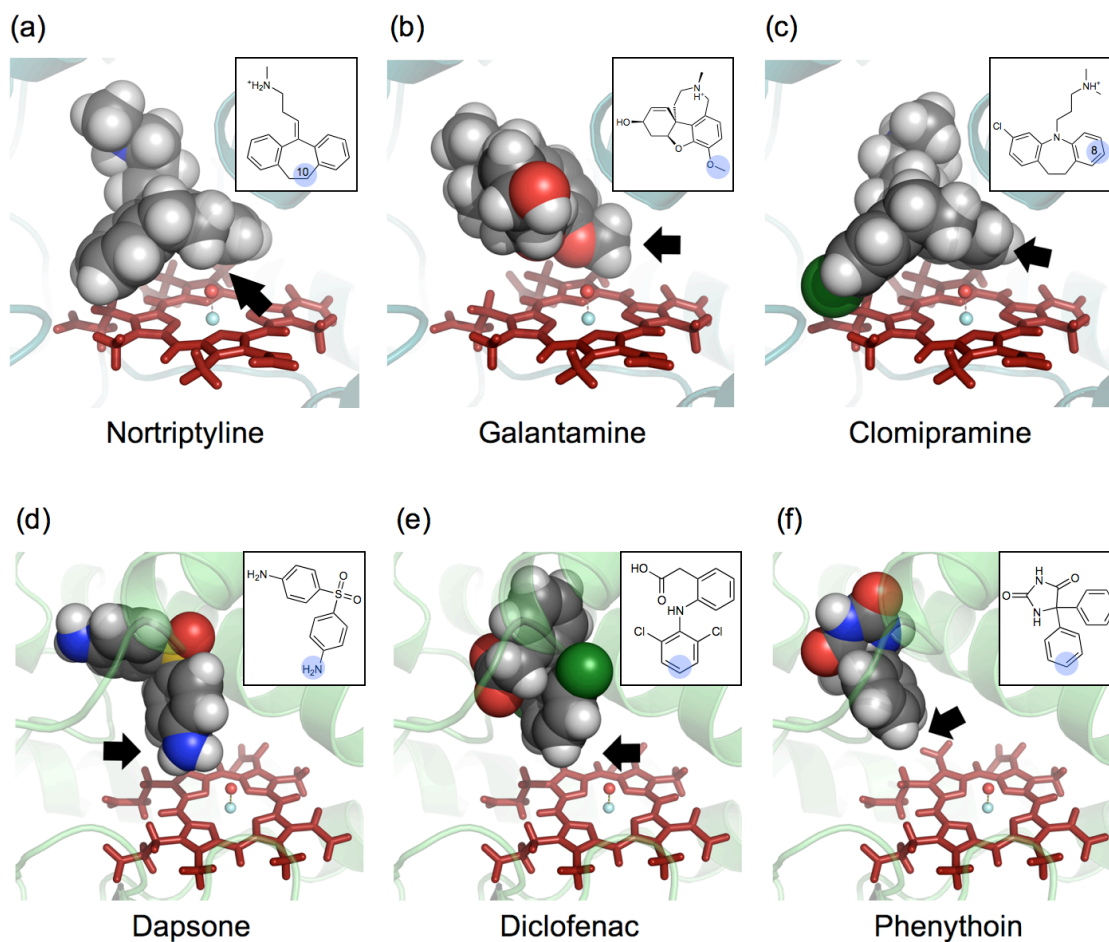


Figure 34: Docking poses as obtained from pre-alignment and automated, flexible docking. a–c: nortriptyline, galantamine and clomipramine binding to CYP2D6, d–f: dapsone, diclofenac and phenytoin binding to CYP2C9. The secondary structure of the enzyme is represented by cartoons, ligands in space filling mode, the heme portion by red sticks, iron and oxygen as spheres. Black arrows indicate the primary sites of metabolism as obtained from experimental data.

2.5 Y-scrambling and binding of external datasets to CYP2D6 and CYP2C9

To investigate the sensitivity and to further scan the predictive capability both towards binding and non-binding compounds, the two models were subject to Y-scrambling. An ensemble of 79 compounds for CYP2D6 (15 external binders and 64 non-binders) and of 96 compounds for CYP2C9 (40 external binders and 56 non-binders) were tested by the respective models to check the relative deviance from the known experimental binding affinities.

The Y-scrambling technique²⁰⁰ shuffles the experimental binding affinities data on which the model is built. If the prediction of the test set reproduces a good correlation using this “random” model, this is an indication that the sensitivity towards the biological data is poor. Forty Y-scrambling tests (using a different random-number seed) were performed for each target, utilizing the same settings used for the “normal” models. Out of the 40 models, the average predictive r^2 is of -0.318 for CYP2D6 and -0.283 for CYP2C9 (Table 5 and 6, Appendix 2 – Supporting Information), demonstrating the sensitivity of the models towards the biological data.

When QSAR models are established, it is important to avoid the overfitting of these models. To correctly handle this problem, the evolution of predictive r^2 and cross-validated r^2 were monitored every 10 generations, to subsequently interrupt the evolution of the genetic algorithm when the predictive r^2 was starting to drop while the cross-validated r^2 was continuing to raise.

In addition, the binding of six classic antihistamines²⁰¹ and nine compounds from the Strobl group²⁰² within the same binding range as the training set were simulated. In this simulation 60% of the ligands are predicted within a factor of 10 in K_i from the experimental data, 86.7% within a factor of 25 and 13.3% with more than a factor of 25 (Table 15).

In a similar way, we have simulated 40 different inhibitors binding to CYP2C9²⁰³ within the same range of binding affinities used to build up the model. In this instance, 80% of the ligands are predicted within a factor of 10 in K_i from the experimental data, 87.5% within a factor of 25 and 12.5% differ more than a factor of 25 (Table 16).

Table 15: External dataset of compounds binding to CYP2C9. Compound structure, compound name, experimental (exp.) and calculated (calc.) binding affinity, and in bold the *Factor off* of the predicted compared to the experimental value (calc/exp).

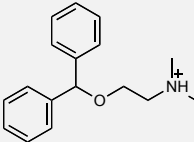
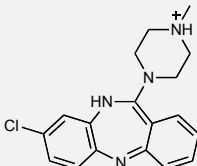
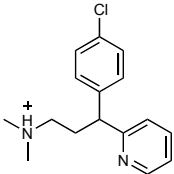
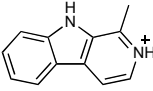
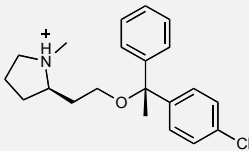
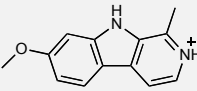
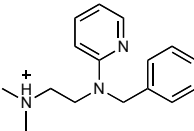
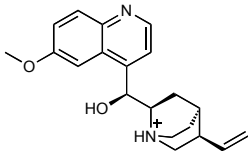
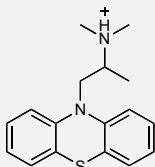
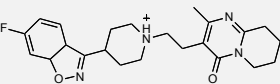
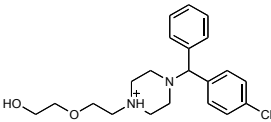
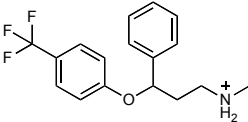
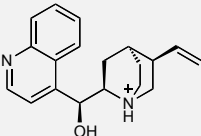
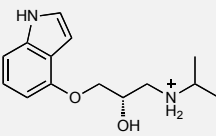
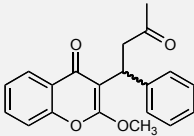
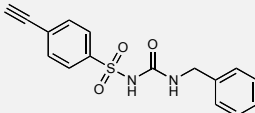
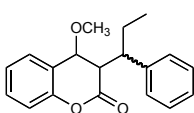
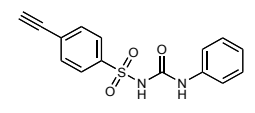
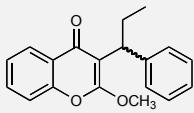
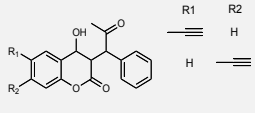
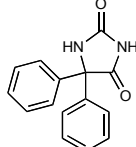
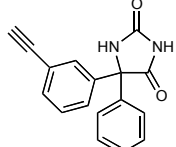
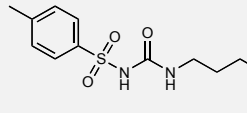
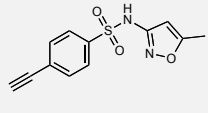
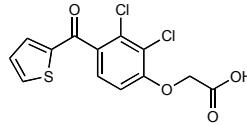
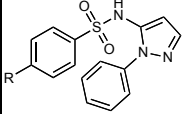
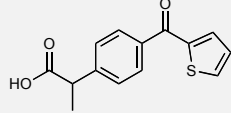
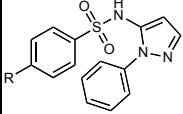
Compound Structure	Exp / Calc K_j Factor_{off} (calc/exp)	Compound Structure	Exp / Calc K_j Factor_{off} (calc/exp)
	Diphenhydramine 11.7 μM exp 28.8 μM calc F_{off} = 2.5		Clozapine 19.0 μM exp 33.2 μM calc F_{off} = 1.7
	Chlorpheniramine 10.8 μM exp 25.9 μM calc F_{off} = 2.4		Harmane 85.6 μM exp 960.2 μM calc F_{off} = 11.2
	Clemastine 2.0 μM exp 44.4 μM calc F_{off} = 22.2		Harmine 50.0 μM exp 522.8 μM calc F_{off} = 10.4
	Tripeleminamine 5.6 μM exp 81.5 μM calc F_{off} = 14.6		Quinidine 4.6 μM exp 156.4 μM calc F_{off} = 34
	Promethazine 3.8 μM exp 11.8 μM calc F_{off} = 3.1		Risperidone 21.4 μM exp 5.3 μM calc F_{off} = 0.2
	Hydroxyzine 3.9 μM exp 1.6 μM calc F_{off} = 0.4		<i>R</i> -Fluoxetine <i>S</i> -Fluoxetine 1.4 μM exp 0.2 μM exp 3.1 μM calc 178.5 μM calc F_{off} = 2.2 F_{off} = 892.5
	Cinchonine 3.5 μM exp 71.9 μM calc F_{off} = 20.5		Pindolol 4.8 μM exp 11.2 μM calc F_{off} = 2.3

Table 16: External dataset of compounds binding to CYP2C9. Compound structure, compound name, experimental (exp.) and calculated (calc.) binding affinity, and in bold the *Factor off* of the predicted compared to the experimental value (calc/exp).

Compound Structure	Exp / Calc K_j Factor_{off} (calc/exp)	Compound Structure	Exp / Calc K_j Factor_{off} (calc/exp)
	E1-9R 31.4 μM exp 7 μM calc F_{off} = 0.2		E14 15 μM exp 25.8 μM calc F_{off} = 1.7
	E1-9S 5.9 μM exp. 4 μM calc F_{off} = 0.7		E15 50 μM exp 193.0 μM calc F_{off} = 3.9
	E-(9)R(11)R 7 μM exp 36.0 μM calc F_{off} = 5.1		E2-(9)S(11)S 12 μM exp. 88.5 μM calc F_{off} = 7.4
	E2-(9)R(11)S 20 μM exp 79.3 μM calc F_{off} = 4.0		E2-(9)S(11)R 10 μM exp 155.3 μM calc F_{off} = 15.5
	E3 0.6 μM exp 6.9 μM calc F_{off} = 11.5		E16 2 μM exp 181.3 μM calc F_{off} = 90.7
	E4-(9)R 0.6 μM exp 10.5 μM calc F_{off} = 17.5		E4-(9)S 0.5 μM exp 39.9 μM calc F_{off} = 80.0
	E5-(9)R(11)R 37 μM exp 11.2 μM calc F_{off} = 0.3		E5-(9)S(11)S 1 μM exp 3.4 μM calc F_{off} = 3.4
	E5-(9)R(11)S 14 μM exp 1.4 μM calc F_{off} = 0.1		E5-(9)S(11)R 0.1 μM exp 31.3 μM calc F_{off} = 313.0
	E6-(9)R 10 μM exp 12.0 μM calc F_{off} = 1.2		E6-(9)S 11 μM exp 15.0 μM calc F_{off} = 1.4
	E17 0.2 μM exp 10.3 μM calc F_{off} = 51.5		E18 18 μM exp 6.8 μM calc F_{off} = 0.4
	E19 48 μM exp 8.9 μM calc F_{off} = 0.2		

 <p>E7-(9)<i>R</i> 29 μM exp 10.3 μM calc F_{off} = 0.4</p>	 <p>E20 25 μM exp 14.6 μM calc F_{off} = 0.6</p>
 <p>E8-(9)<i>R</i> 13 μM exp 87.6 μM calc F_{off} = 6.7</p>	 <p>E21 7 μM exp 25.1 μM calc F_{off} = 3.6</p>
 <p>E9-(9)<i>R</i> 14 μM exp 59.7 μM calc F_{off} = 4.3</p>	 <p>E22 1.5 μM exp 12.5 μM calc F_{off} = 1.7</p> <p>E23 1.1 μM exp 5.5 μM calc F_{off} = 5.0</p>
 <p>E10 10 μM exp 26.0 μM calc F_{off} = 2.6</p>	 <p>E24 18 μM exp 46.3 μM calc F_{off} = 2.6</p>
 <p>E11 70 μM exp 28.9 μM calc F_{off} = 0.4</p>	 <p>E25 3 μM exp 26.2 μM calc F_{off} = 8.7</p>
 <p>E12 11 μM exp 11.3 μM calc F_{off} = 1.0</p>	 <p>E28 2.4 μM exp 0.3 μM calc F_{off} = 0.1</p> <p>E29 2 μM exp 8.7 μM calc F_{off} = 4.4</p>
 <p>E13 45 μM exp 94.8 μM calc F_{off} = 2.1</p>	 <p>R: 1) CH₃-CO- 2) H- 3) OH- 4) CH₃</p> <p>E30 6 μM exp 5.5 μM calc F_{off} = 0.9</p> <p>E31 0.1 μM exp 7.7 μM calc F_{off} = 77</p>

The data used to build the model originate from a filtered Merck repository¹⁰⁵ of drug-like compounds where the majority of the compounds present affinities equal or higher than 200 μM either to CYP2D6 or to CYP2C9.

To further challenge our model and to evaluate the probability to obtain “false positive” affinity values for low-affinity or not binding compounds, we simulated the binding of 64 non-binders to CYP2D6 and 56 of non-binders to CYP2C9. The simulation were performed using the protocol implement in the *VirtualToxLab*,¹⁰⁹ where the ligands pass through an extensive (5,000 steps) conformational search in aqueous solution, protonation and tautomeric state calculation at physiological pH followed by pre-alignment of the low energies conformation of the ligands, flexible docking and binding affinity prediction using mQSAR.

For CYP2D6, 56 out of 64 (85.9%) and for CYP2C9, 46 out of 56 (82.1%) ligands were predicted to be non-binders or to bind within a factor of 10 in K_i from the weakest binder of the training set (praziquantel for 2D6 and protriptyline for 2C9) (Figure 35).

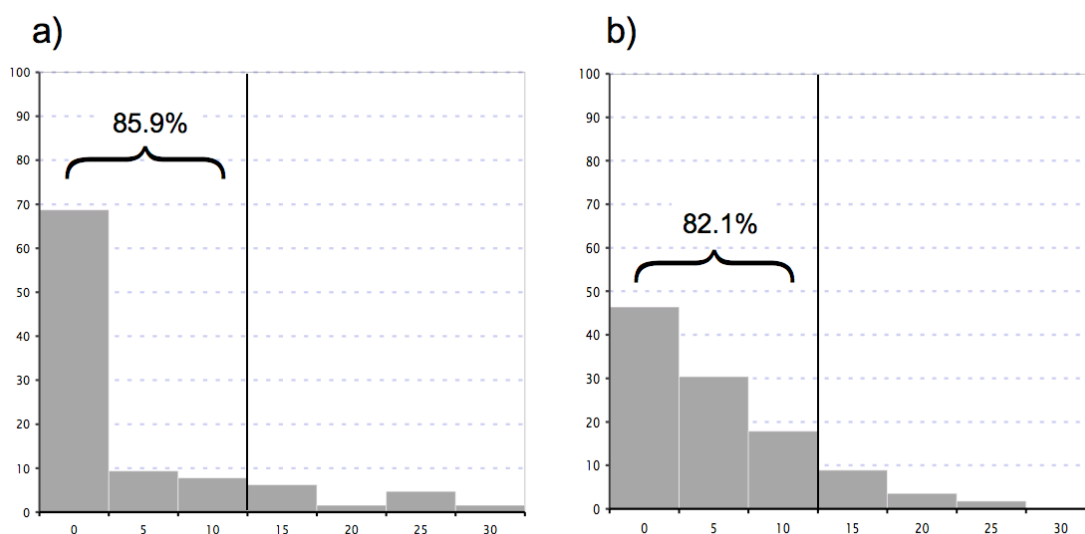


Figure 35: Plotted percentages of compounds (y-axis) and error factor in K_i (x-axis) from the weakest binder of CYP2D6 (a) and to CYP2C9 (b). The vertical line isolates the compounds predicted to have a K_i within a factor of 10 from the weakest binder of the two datasets.

2.6 VirtualToxLab dataset screening

The models presented herein for CYP2D6 and CYP2C9 were added to the *VirtualToxLab*, an *in-silico* tool for predicting the toxic potential (endocrine and metabolic disruption) of drugs, chemicals and natural products. Through automated flexible docking combined with mQSAR it simulates and quantifies the interactions of drug-like compounds to 16 different proteins (the androgen,²⁰⁴ aryl hydrocarbon,²⁰⁵ estrogen α/β ,¹⁶⁷ glucocorticoid,²⁰⁶ liver X,²⁰⁷ mineralocorticoid,²⁰⁸ peroxisome proliferator-activated receptor γ ,²⁰⁹ thyroid α/β receptors,²¹⁰ the enzymes CYP1A2, CYP2A13, CYP2C9, CYP2D6, CYP3A4,¹⁶⁰ as well as the hERG ion channel).

The VTL repository of 2,518 drugs, chemicals and natural products have been screened at the Biographics Laboratory 3R and the predicted toxicity alerts are listed at the website <http://www.virtualtoxlab.org>.

It should be noted that amphetamine and methylenedioxyamphetamine are predicted to bind to CYP2D6 with a K_i of 8.9 μM and 1.6 μM . In literature, both methamphetamine and methylenedioxymethamphetamine are metabolized by CYP2D6.²¹¹ Similarly, sertraline displays a potential in mediating drug-drug interactions as well as inhibition of CYP2D6.²¹² Our totally automated protocol predict sertraline as a high-affinity binder to CYP2D6 with a K_i of 4.1 μM .

A recent survey reported pantoprazole as one of the most potent CYP2C9 in vitro inhibitors among the currently used proton pump inhibitors (PPI) with a K_i value of 6.5 μM .²¹³ The screening of the VTL repository on CYP2C9 scored pantoprazole as one of the most affine binder to CYP2C9 with a predicted K_i of 33.1 μM , indicating an appreciable small difference between experimental and predicted affinity (a factor of five).

3 Computer-aided drug design at the Myelin Associated Glycoprotein (MAG)

3.1 Experimental data and molecular docking

In Table 17, 18, 19 and 20 are presented binding affinities for different glycomimetics. The text is organized in *Series (1, 2, 3, and 4)*. The results obtained for *Series 1*, and 2 are part of publicly available scientific articles and they are attached in Appendix 3 and 4. The results for *Series 3* and 4 are part of scientific articles not yet published.

3.1.1 Series 1

The binding affinities (rIC_{50} , K_D) and LogD value of compounds A1, A2, and B1 are presented in Table 17. The techniques used to obtain the experimental affinities are mentioned and described in detail in the methods section of Appendix 3 and 4.

Table 17: A and B represent the two core moieties substituted in position R1, R2, and R3 (A1, A2, B1). In the table are listed the relative inhibitory concentrations rIC_{50} (using compound A1 as reference), K_D and $\log D_{7.3}$ values.

A

B

#	R1	R2	R3	rIC_{50}	K_D [μM]	$\log D$ (pH 7.3)
A2			Cl	0.02	0.5	-0.26
B1			Cl	4300	>800	n.d

Considering compounds **A1**, **A2** and **B1** (Table 17), two slightly different docking modes are devised especially when comparing compound A2 and B1 (Figure 36). In both the compounds, the anchor point of Arg 118 was used to interactively dock the ligands. In addition, the hydrogen atom of the 5-NHFAC groups interacts with the backbone carbonyl of Gln126, and the C(4)-OH (in A2) or C(4)-NH (in B1), respectively, establishes hydrogen bonds with the side chain carbonyl of Gln126. In A2, the 2,3-difluorobenzyl aglycone and in B1 the benzyloxy substituent are located within a hydrophobic pocket formed by Trp59, Tyr60, Tyr69, and Tyr116. However, while in A2 the C(9)-NH of the benzamido group establishes a H-bond with the backbone carbonyl oxygen atom of Thr128, in B1 the 9-benzamido carbonyl interacts with the backbone NH of Thr128, thus forming a much weaker H-bond than the parent compound A2. Consequently, the orientation of the glycerol side chain in A2 is different compared to the 2-benzamido-ethyl side chain in B1. This modification reduces the interaction of the p-chlorobenzamide in B1 with the second hydrophobic pocket lined by Ser130 and Glu131.

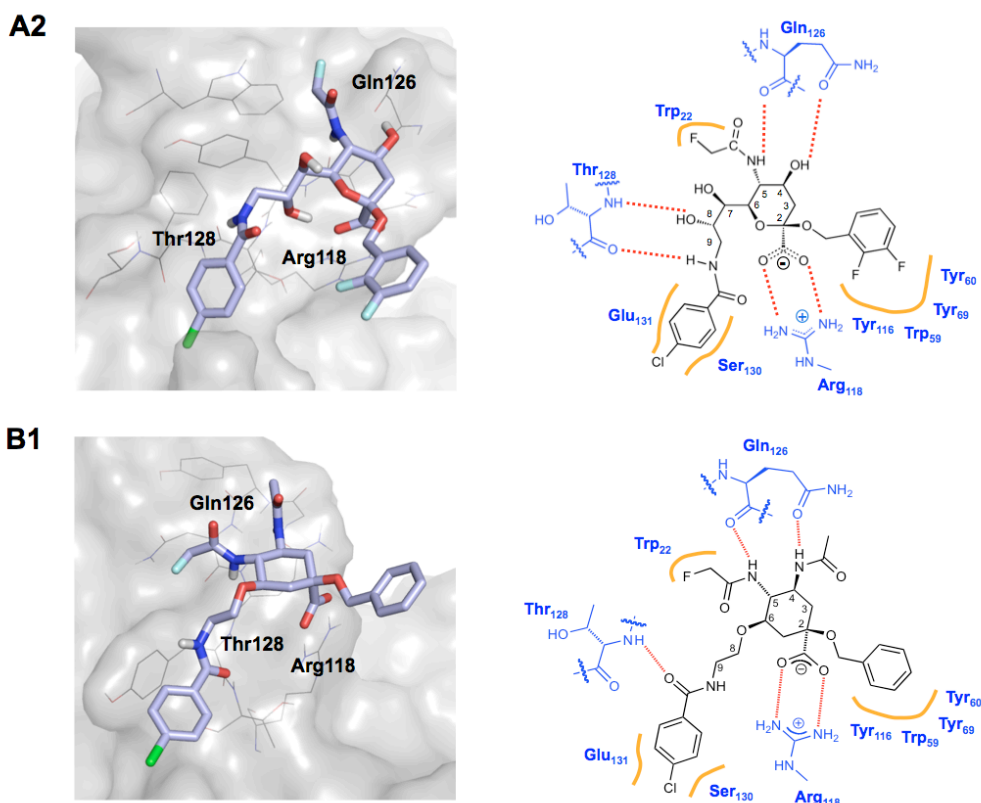


Figure 36: Docking modes (left) for compounds A2 and B1. Ligands are shown colored in cyan and in sticks representation. MAG is depicted as molecular surface. The main interactions involved in the binding are shown on the right with hydrogen bonds represented as red lines.

3.1.2 Series 2

Similarly to A2 and B1, the compounds A3–A18 (Table 18), were docked using the salt bridge between the carboxylic acid and Arg118 as main anchoring point.

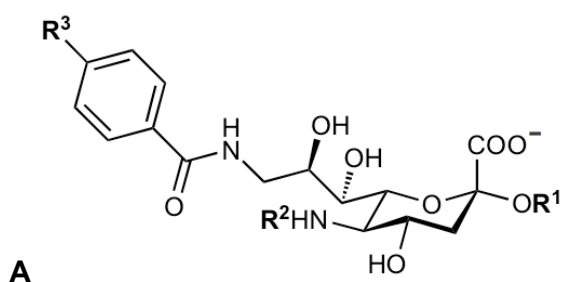
In Figure 36 – A2 are visible additional hydrogen-bonds between the 5-NH and the backbone carbonyl of Gln126, the carboxylate and the OH of Thr128, the 8-OH and the backbone NH of Thr128 and the 9-NH and the backbone carbonyl of Thr128. This latter finding is in good agreement with previous studies, where the lack of a hydrogen bond donor at position 9 resulted in a reduced binding affinity.¹³⁷ A considerable contribution to the binding affinity results from hydrophobic interactions. Thus, the *p*-chlorobenzamide is shown to point into a hydrophobic pocket, built by Ser130 and Glu131. The second hydrophobic pocket, which hosts the aglycone substituents, is defined by the side chains of Trp59, Tyr60, and Tyr69.



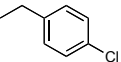
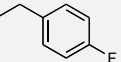
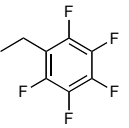
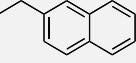
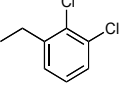
With respect to different substitution patterns, the dichloro compound A10 shows a 4-fold and the difluoro compound A11, a 7-fold enhancement in affinity compared to the reference compound A4, indicating a charge transfer complex with the electron rich aromatic ring of Tyr60. The only moderate improvement in binding affinity for the halogenated compounds A6 and A7 could be due to a steric clash of the *p*-substituent with the protein. Compound A8 was synthesized as a symmetric analog of compound A11 in order to compensate entropy loss due to the orientation of the 2,3-difluorobenzyl ring. Again, there is no improvement of the binding affinity in comparison to compound A11, which may be the consequence of a sterically too bulky substituent in the *para* position. Finally, the improved affinity of A9 might originate from favorable π - π interactions of the naphthyl with Tyr69.

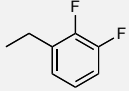
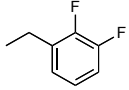

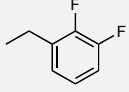

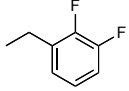
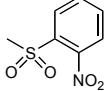
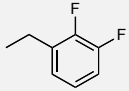

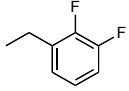
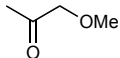
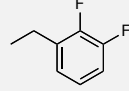
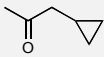
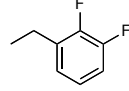
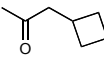
Some of the compounds modified at the 5-position seem to form a favorable σ - π interaction with Trp22. In case of A12 and A13, we assume that the positively polarized hydrogens of the FAc or ClAc, respectively, are oriented towards the aromatic ring.²¹⁴ As fluorine is more electronegative than chlorine, the polarization of the hydrogens is stronger and therefore the interaction is more favorable. For compounds A17 and A18, additional hydrophobic interactions are observed, but the binding site seems to be spatially limited.²¹⁵ The reduced affinity of A15 could be a consequence of the different bond angle for the sulfonamide substituent compared to an acetate in the same position (A11), leading to different spatial requirements. While methylsulfonamide A15 shows a decrease in binding affinity, the nosyl substituent (A14) shows an opposite behavior. This might be due to the formation of a charge-transfer complex with

Trp22. To summarize, modifications at the reducing end improved binding affinity by a factor of 7 (A4→A12). Combined with the best modification at the 5-position, a high affinity ligand A12 was obtained.

Table 18 (continued on the next page): A represents the core moiety substituted in position R1, R2, and R3 (A3–A18). In the table the relative inhibitory concentrations rIC_{50} (using as reference compound A4), K_D and $\log D_{7.3}$ values are listed.



#	R1	R2	R3	rIC_{50}	K_D [μ M]	$\log D$ (pH 7.3)
A3	-CH ₃	Ac	H	8.00	137	n.d.
A4		Ac	Cl	1.00	17	n.d.
A5		Ac	H	1.50	26	n.d.
A6		Ac	Cl	1.30	15	0.36
A7		Ac	Cl	1.20	13	-0.11
A8		Ac	Cl	0.26	6.1	-0.11
A9		Ac	Cl	0.74	11.6	0.58
A10		Ac	Cl	0.50	4.3	0.53

A11		Ac	Cl	0.30	2.4	-0.27
A12			Cl	0.02	0.5	-0.26
A13			Cl	0.07	2.1	0.35
A14			Cl	0.05	1.4	0.87
A15			Cl	0.60	17	-0.17
A16			Cl	0.14	2.3	0.06
A17			Cl	0.10	4.1	0.31
A18			Cl	0.14	5.8	0.75

3.1.3 Series 3

A further series of mimetics having the moiety C as starting core were docked to the MAG model aiming to improve the understanding of their binding affinity (Table 19).

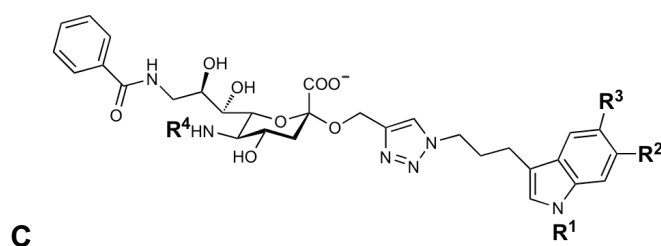
In both antagonist-protein complexes, the sialic acid core established the crucial interactions identified in the previous docking studies and responsible for recognition and binding: the salt bridge between Arg118 and the carboxylate of the ligand and the hydrogen bonds between 5-NH and the carbonyl of Gln126, 8-OH and Thr128 and 9-NH and Tyr125.



In the case of C9, the 5'-nitroindole is embedded in a hydrophobic pocket lined by residues Tyr60, Tyr69 and Tyr116. A further electrostatic interaction of the

nitro substituent could be established with Lys 67. With this interaction the indole-moiety could establish a π - π stacking to Tyr69 and 116 (Figure 37 – A). The linker contributes only marginally to the overall binding affinity but enables the proper alignment of the indole. In the case of the nitro-indole substituent, the linker seems to be in a rather locked conformation after binding (Figure 37 – A) compared to the unsubstituted indole (Figure 37 – B).

In the case of the unsubstituted indole (Compound C4, Figure 37 – B) no π - π interactions were observed. These interactions were analyzed monitoring the orientation of the plane of the indole ring towards the two planes of Tyr69 and Tyr116. The angles relative to the interactions of the indole with the tyrosines are both around 70°-90° and the dispersion is clearly higher.

Table 19: With C is represented the core moiety substituted in position R1, R2, R3, and R4 (C1–C9), along with the correspondent K_D values.



Compound	R ¹	R ²	R ³	R ⁴	K _D [nM]
C1	H	H	NO ₂	Ac	190-397
C2	Me	H	NO ₂	Ac	570
C3	Et	H	NO ₂	Ac	390
C4	H	H	H	Ac	320
C5	H	H	OMe	Ac	370
C6	H	H	CF ₃	Ac	510
C7	H		H	Ac	480
C8	H		H	Ac	170-75
C9	H	H	NO ₂	FAc	60

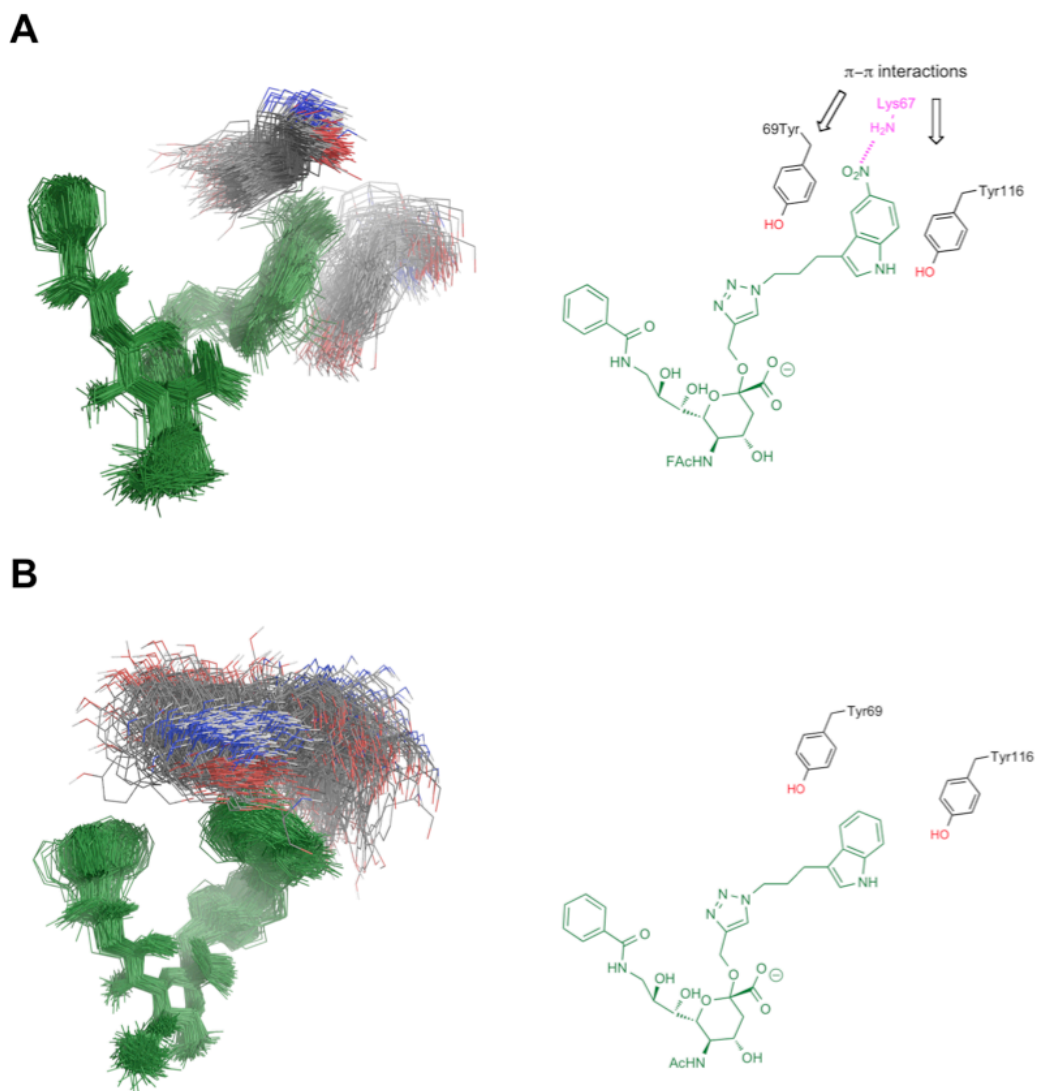
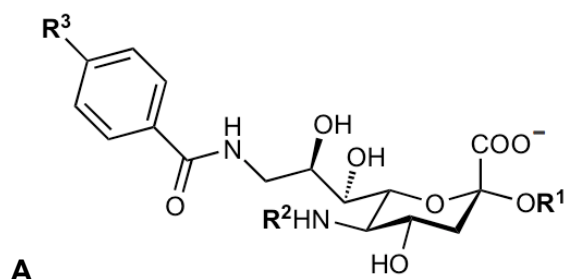


Figure 37: Superposition of the ligand structures C9 (A) and C4 (B), from the most populated cluster of ligand conformations from the MD simulations. On the right two dimensional representation of the interaction with Tyr69 and Tyr116.

Seeing the promising affinities found for compound C9 (Figure 38 – A), and noticing the metabolic instability of the nitro-substituent, which in such an accessible position could be easily metabolized, further substitutions in positions R1 (Table 20, next page) were simulated to design novel potent ligands lacking of this metabolic liability. In Figure 38 are shown the different docking modes of compounds having the nitro group substituted with a cyano group (Figure 38 – B) and a sulfonylmethyl group (Figure 38 – C), respectively.

Table 20: Core moiety A with further substitutions in position R1, R2, and R3



#	R ¹	R ²	R ³
A-19			H
A-20			H
A-21			H

For the three substitutions the interactions of the core moiety with Arg118, Gln126, and Thr128 (Figure 38, next page – interactions 1, 2a–c, 3, 4, 4a, 5) were established and their stability was monitored throughout molecular dynamics simulations (cf. chapter 3.2.3). In all the three docking modes, which correspond also to the starting position of the MD simulations, the nitro, cyano and sulfonylmethyl substituents were located in proximity to Lys67 to allow for a possible interaction.

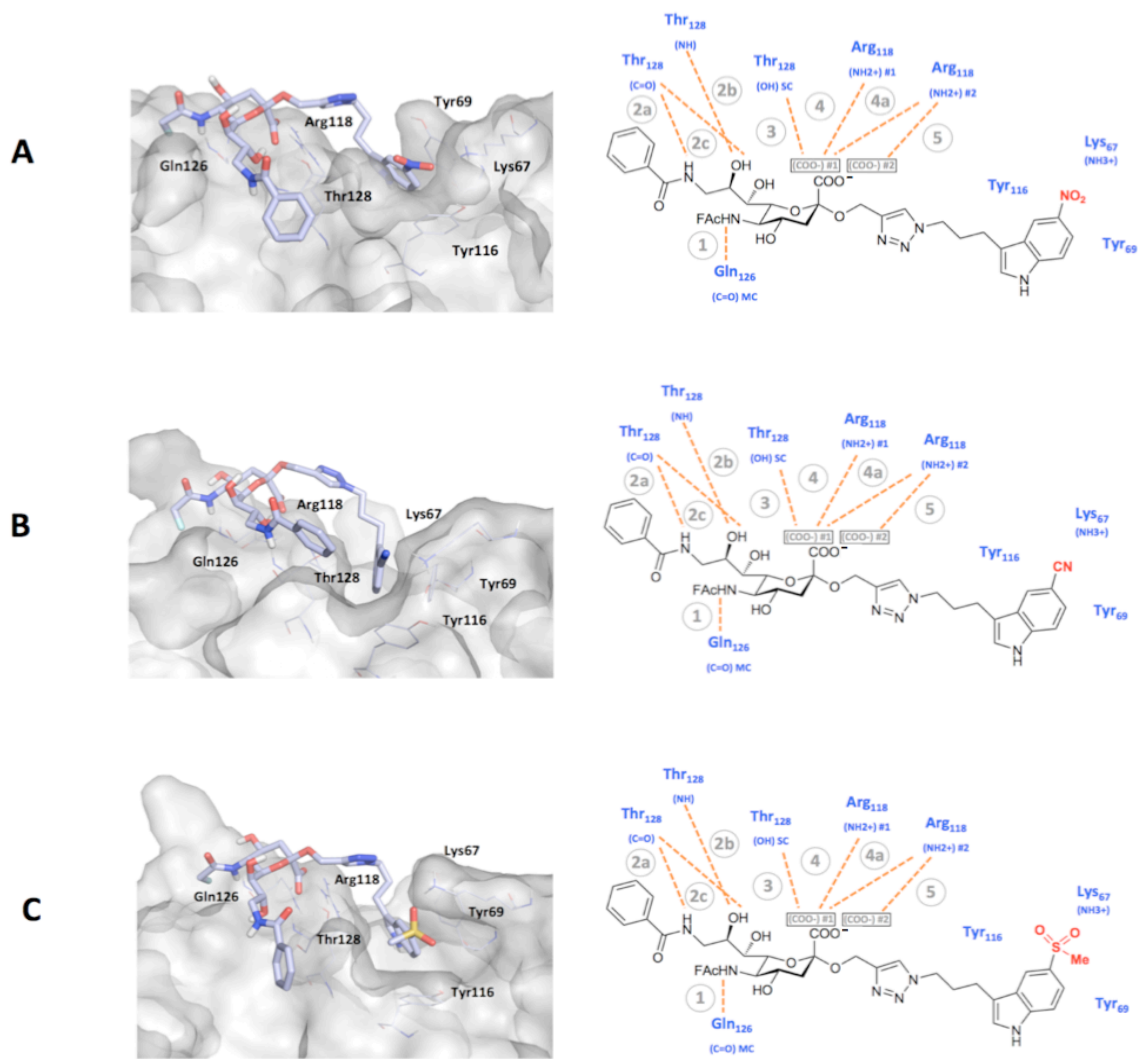
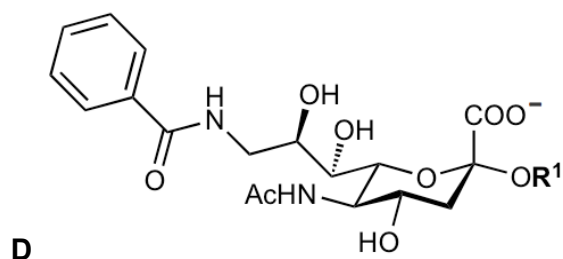


Figure 38: Docking modes of A19 (A), A20 (B), and A21 (C). On the right the main interactions (numbered) involved in ligand binding.

3.1.4 Series 4

The compounds related to the core moiety D (i.e. D1, D2, D3 – Table 21) were docked using the same interactions explained in the previous series (Table 18, 19 and 20 and Figure 36 and 38). Starting from these docking modes, the compounds were analyzed in MD simulations to check if a correlation between STD-NMR signals and MD simulations was identifiable.

Table 21: D represents the core moiety substituted in position R1 (D1–D3), along with the correspondent K_D values.



Compound	R ¹	K _D [μM]
D1		33
D2	Me	131
D3		20

3.2 Molecular-dynamics simulations

Ligand binding should be described more like a dynamic event than like a static representation of the most important interactions involved in binding. A study of the dynamic of ligand binding can improve the understanding of ligand–protein binding events. Compared to rigid molecular docking, MD simulations allow rearrangements of both ligand and protein atoms (for both backbone and side-chains atoms) and they simulate different conformations of the ligand–protein complex over the simulation time. Consequently, MD simulations minimize the possibility of a *local minimum* configuration of the complex, thus they can provide a more comprehensive overview of the kinetics of binding when the sampling time is long enough to be comparable with the real time of binding. During MD simulations it is also possible to evaluate the contribution of each non-bonded interaction over the simulation time.

3.2.1 Series 1

The reason for synthesizing the non-carbohydrate mimic B1 of the high-affinity MAG antagonist A2 was to improve its pharmacokinetic profile. The two compounds were docked to the homology model of MAG forming the intermolecular interactions explained in the previous section. Although B1 can establish all but one of the essential hydrogen bond interactions present in A2, a dramatic drop in affinity was observed (a factor of more than 1600 in K_D).

The two complexes differ in the stability of their hydrogen-bond network. The strength and frequency of the salt bridge to Arg118 is comparable for both A2 and B1 (entries 1 & 2), as well as the hydrogen bonds of the 5-fluoroacetamido NH with the backbone carbonyl of Gln126 (Table 22).

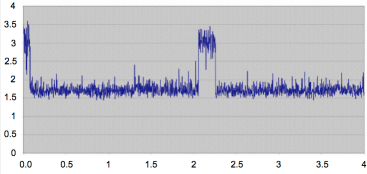
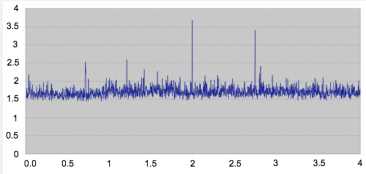
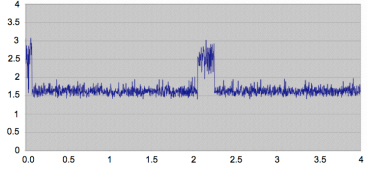
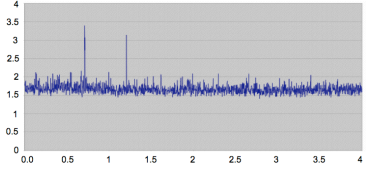
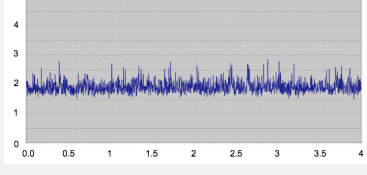
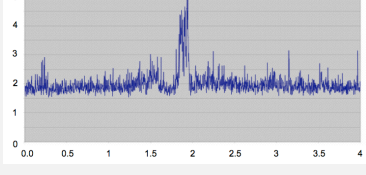
Despite an almost stable interaction of the core moiety, major differences are found for the interaction patterns of the side chains, i.e. the glycerol side chain in A2 and the 2-benzamidoethoxy side chain of B1. In the compound A2, two important hydrogen bonds contribute to binding: (i) the 9-benzamido NH interacts with the backbone carbonyl of Thr128 (1.96 Å, entry 4) and (ii) the backbone NH of the same amino acid acts as donor in a H-bond with the oxygen of the C(8)-OH (1.94 Å, entry 6).

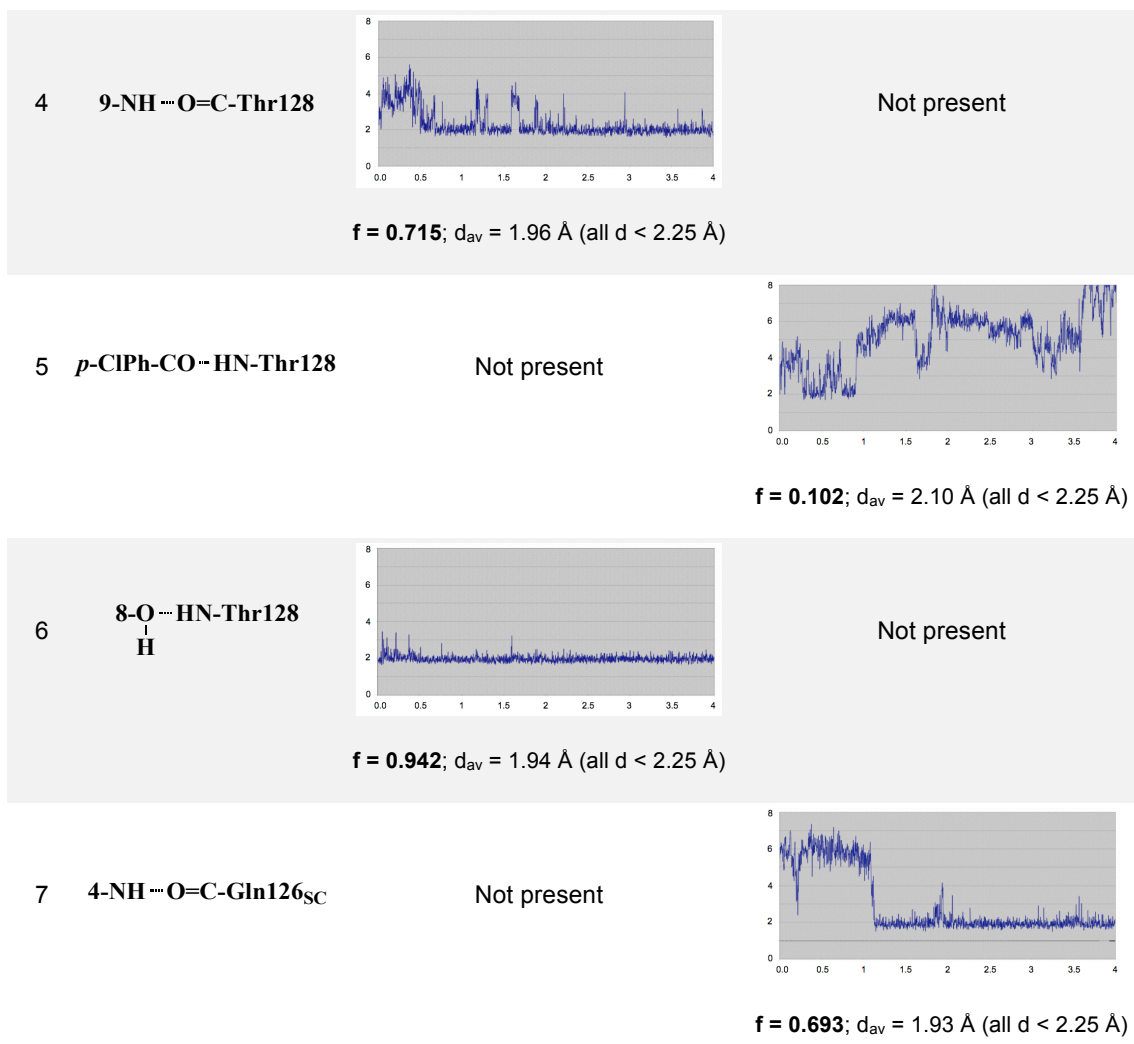
Since the benzamidoethoxy side chain in B1 is oriented in an inverted manner, a significantly weaker H-bond between the backbone NH of Thr128 and the 9-

benzamido carbonyl (2.10 Å, entry 5) is formed. In addition, the hydroxy group at C(8) is absent, therefore the second H-bond is not present at all. As a consequence, the orientation of the side chain alters, leading to a markedly reduced hydrophobic contact of the *p*-chlorobenzamide with Ser130 and Glu131.

Finally, the hydrogen bond of the C(4)-OH in A2 with the side chain carbonyl of Gln126, which was identified in the static docking studies, turned out to be irrelevant in the kinetic considerations. Although the hydrogen bond of the 4-acetamido NH in B1 with the side chain carbonyl of Gln126 substantially contributes to the binding to MAG (1.93 Å, entry 7), unexpectedly this additional interaction does not compensate the experimentally observed loss of affinity caused by the reduced hydrophobic contact of the benzamido group in B1 compared to A2.

Table 22: Time-resolved H-bond distances between key amino-acid residues of MAG and compounds A2 and B1. The horizontal axis represents the simulation time, the vertical axis the H...acceptor separation. *f*: frequency of observance of the H-bond throughout the molecular-dynamics simulation; *d_{av}*: average H...acceptor distance; MC: main chain; SC: side chain.

Entry	Interaction	Compound A2	Compound B1
1	$-\text{COO}^{\ominus} \cdots \text{H}_2\text{N}^{\oplus}\text{-Arg118}$ (pair #1)		
		f = 0.922 ; <i>d_{av}</i> = 1.71 Å (all <i>d</i> < 2.0 Å)	f = 0.973 ; <i>d_{av}</i> = 1.69 Å (all <i>d</i> < 2.0 Å)
2	$-\text{COO}^{\ominus} \cdots \text{H}_2\text{N}^{\oplus}\text{-Arg118}$ (pair #2)		
		f = 0.934 ; <i>d_{av}</i> = 1.62 Å (all <i>d</i> < 2.0 Å)	f = 0.984 ; <i>d_{av}</i> = 1.66 Å (all <i>d</i> < 2.0 Å)
3	5-NH...O=C-Gln126 _{MC}		
		f = 0.945 ; <i>d_{av}</i> = 1.88 Å (all <i>d</i> < 2.25 Å)	f = 0.948 ; <i>d_{av}</i> = 1.93 Å (all <i>d</i> < 2.25 Å)



The docking modes of A2 and B1 to a homology model of MAG revealed a different orientation of the side chains in the 6-position. Obviously, the C(8)-OH and the C(9)-NH in sialoside A2 represent key polar groups, i.e. the recognition of both of these hydrogen bond donors is required for binding. Consequently, in the absence of the stabilizing H-bond between the C(8)-OH in A2 and the backbone NH of Thr128, the benzamidoethoxy side chain adopts an inverted orientation, leading to a less favorable interaction of its benzamido group with the hydrophobic pocket formed by Ser130 and Glu131 and therefore to the experimentally observed decrease in affinity.

3.2.2 Series 2

Similarly to Series 1, frequency and mean distance of the interactions established between antagonist and MAG were monitored. In Table 23 are shown statistical data on the interactions involved in the binding of A12 and A15 to MAG.

Table 23: Time-dependent hydrogen-bond analysis of the MAG complexes with A12 and A15

	-COO⁻...⁺H₂N-Arg118	>NH ... O=< Thr109	-OH ... HN< Thr109	>NH ... O=< Gln107
A12	f = 0.933 d _{ave} = 1.72/1.64 Å	f = 0.786 d _{ave} = 1.99 Å	f = 0.989 d _{ave} = 1.96 Å	f = 0.984 d _{ave} = 1.90 Å
A15	f = 0.981 d _{ave} = 1.68/1.69 Å	f = 0.090 d _{ave} = 2.16 Å	f = 0.239 d _{ave} = 2.12 Å	f = 0.872 d _{ave} = 1.92 Å

f = frequency of observance throughout the molecular-dynamics simulation;
d_{ave}: average H...Acceptor distance.

While the salt bridge to Arg118 is comparable both in frequency and strength with both A12 and A15, the three hydrogen bonds with residues Gln107 and Thr109 are stronger for A12. The interaction of A15 with Thr109 is less frequent and with a higher average distance when compared to the same interaction to A12. This, could be a possible explanation of the experimental difference in the binding affinity of the two compounds (a factor of 34 in K_D — corresponding to 2.0 kcal/ mol).

3.2.3 Series 3

Frequency and average distance of the interactions involved in the binding of compound A19, A20 and A21 to MAG are shown in Table 24. These highlight a substantial stability and cooperativity of the interactions involved in the binding of the core moiety of these compounds. For compound A19, A20, and A21 the relative orientation of the substituted indole ring of the three compounds towards the aromatic rings of Tyr69 and Tyr116 was investigated (Figure 39). In all three cases the π - π interaction of the indole ring with Tyr69 is relevant where the angle of interaction is generally close to 0° , meaning a parallel positioning of the two rings. In the case of the interaction with Tyr 116 the preferred interaction for the three compounds generally varies between 15° and 60° for compound A19 (Figure 39 – A), 40° and 90° for compound A20 (Figure 39 – B), highlighting a preferred orthogonal positioning of the indole ring towards the aromatic ring of Tyr 116, and between 20° and 70° for compound A21 (Figure 39 – C).

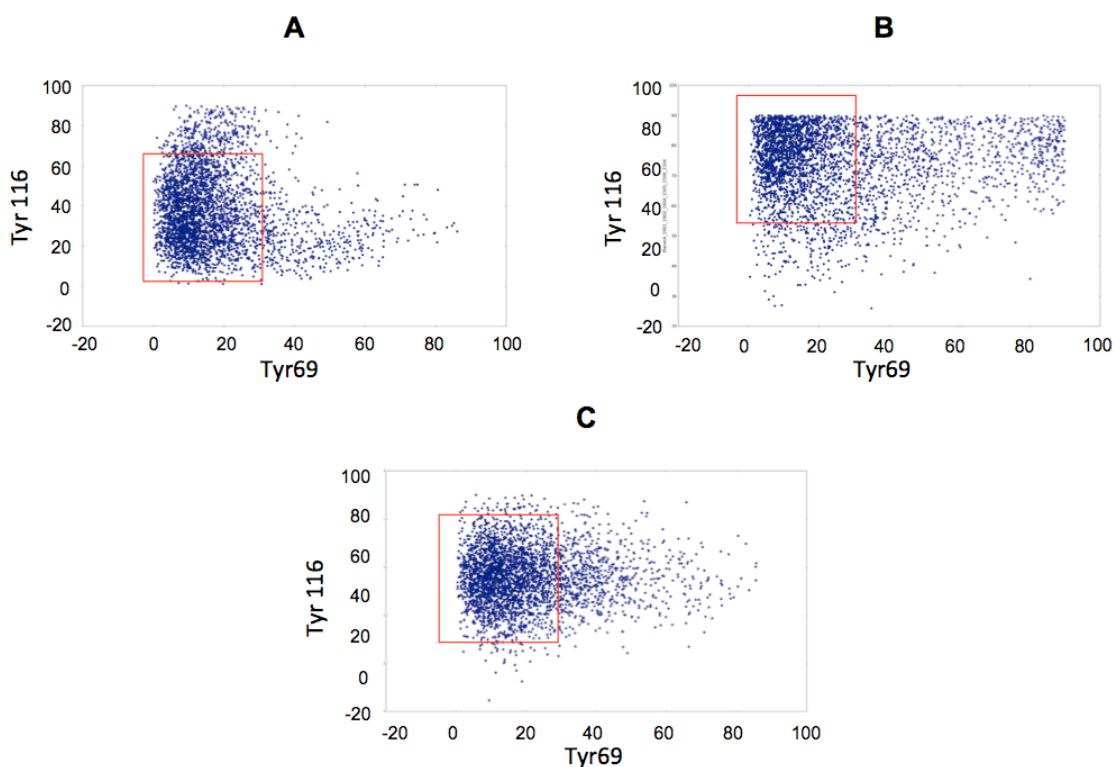
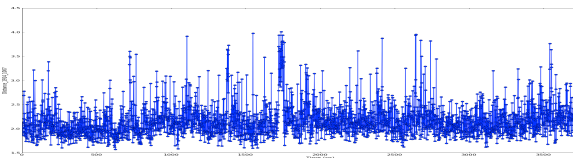
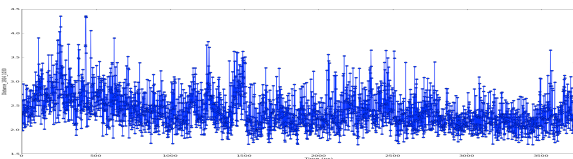
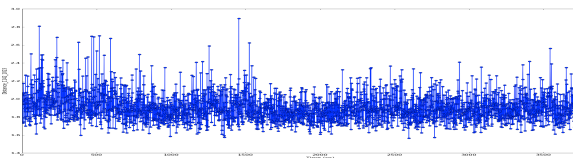
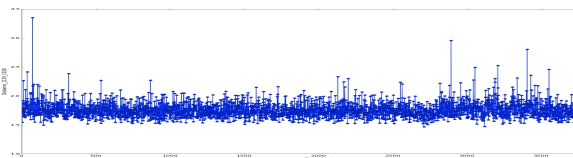
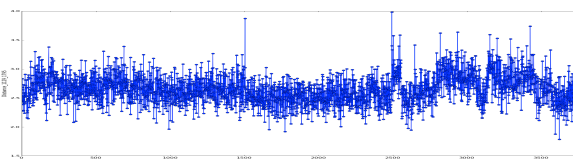
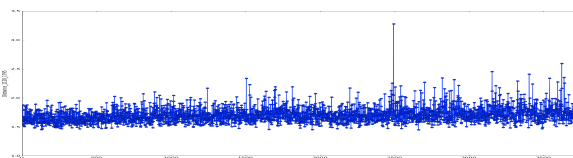
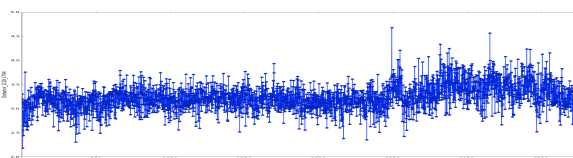


Figure 39: Analysis of the angle between the plane of the indole ring and the plane of Tyr69 and Tyr116, for each of the trajectory frame of the MD simulation of compound A19 (A), A20 (B) and A21 (C). The most populated cluster of angles are highlighted by a red square.

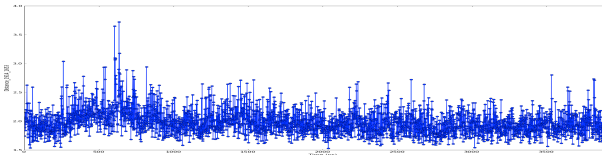
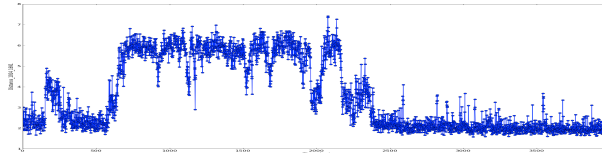
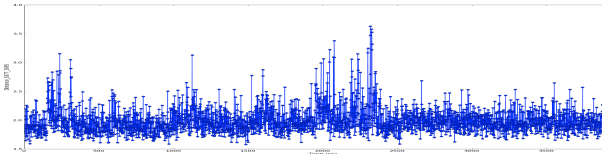
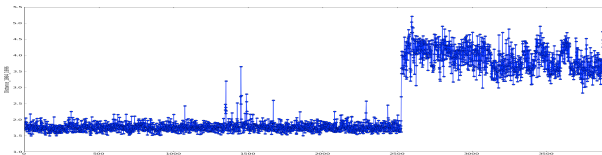
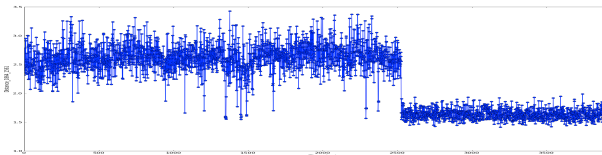
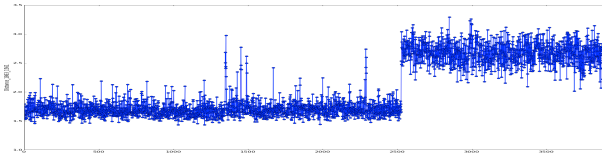
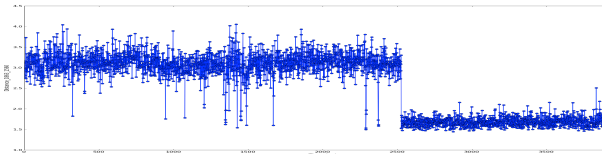
Finally, the interactions of the three indole substituents with the side chain of Lys 67 were monitored. In the case of the nitro substituent (compound A19) an interaction within 2.5 Å between the nitro and the charged NH_3^+ of the lysine is present for 3.12% of the simulation. For the cyano group substituent (compound A20) such an interaction is constituted for 0.24% of the simulation. For the sulfonylmethyl substituent (compound A21) the frequency of interaction is substantially higher (31.44%).

Table 24 (next page): Time-resolved H-bond distances (along with mean, median and standard deviation) between key amino-acid residues of MAG (see Figure 38) and compounds A19, A20 and A21. The horizontal axis represents the simulation time, the vertical axis the H...acceptor separation.

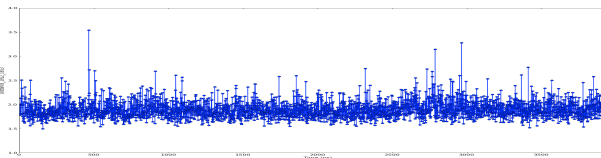
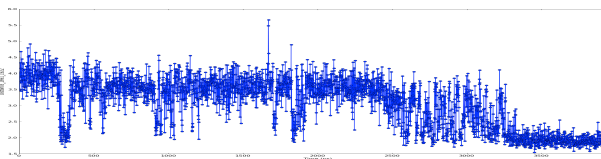
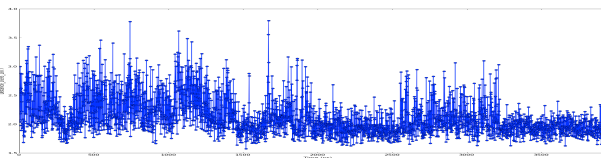
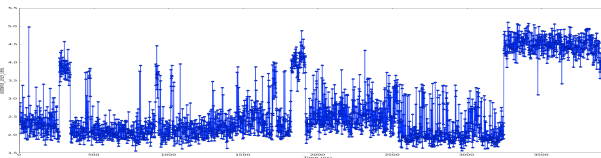
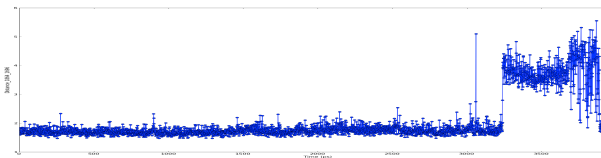
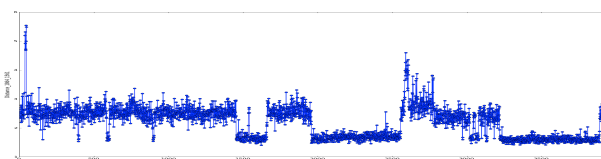
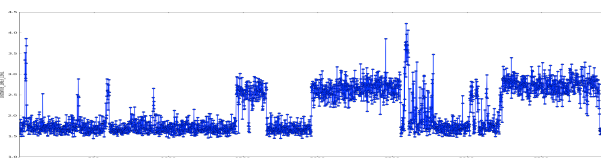
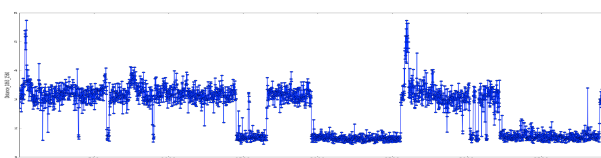
A19: NO₂-Indole substituent

Interaction	Distance (Å)	Time resolved hydrogen-bond distance
1)	Mean: 2.17 Median: 2.10 SD: 0.34	
2a)	Mean: 2.40 Median: 2.33 SD: 0.38	
2b)	Mean: 1.90 Median: 1.89 SD: 0.15	
2c)	Mean: – Median: – SD: –	Not present
3)	Mean: 1.75 Median: 1.74 SD: 0.13	
4)	Mean: 2.64 Median: 2.63 SD: 0.25	
4a)	Mean: 1.71 Median: 1.70 SD: 0.13	
5)	Mean: 3.23 Median: 3.22 SD: 0.27	

A20: CN–Indole substituent

Interaction	Distance (Å)	Time resolved hydrogen-bond distance
1)	Mean: 1.98 Median: 1.95 SD: 0.22	
2a)	Mean: 3.56 Median: 2.71 SD: 1.73	
2b)	Mean: 1.99 Median: 1.95 SD: 0.22	
2c)	Mean: – Median: – SD: –	Not present
3)	Mean: 2.56 Median: 1.86 SD: 1.10	
4)	Mean: 2.27 Median: 2.50 SD: 0.52	
4a)	Mean: 2.10 Median: 1.78 SD: 0.50	
5)	Mean: 2.60 Median: 2.93 SD: 0.72	

A21: SO₂Me–Indole substituent

Interaction	Distance (Å)	Time resolved hydrogen-bond distance
1)	Mean: 1.90 Median: 1.87 SD: 0.17	
2a)	Mean: 3.06 Median: 3.30 SD: 0.78	
2b)	Mean: 2.12 Median: 2.03 SD: 0.32	
2c)	Mean: 2.77 Median: 2.34 SD: 0.95	
3)	Mean: 2.10 Median: 1.75 SD: 0.81	
4)	Mean: 2.20 Median: 2.33 SD: 0.52	
4a)	Mean: 2.11 Median: 1.84 SD: 0.51	
5)	Mean: 2.61 Median: 2.94 SD: 0.80	

3.2.4 Series 4

The aim of the STD-NMR experiments performed by Dr. Brian Cutting (results of the STD-NMR signal for the *N*-benzoate moiety are shown in Figure 40) was to investigate the hydrophobic contact of the *N*-benzoate moiety of compounds D1, D2 and D3 to the protein (Table 25). The relative contribution of the *ortho*, *meta*, and *para* positions of the benzoate was analyzed from the STD-NMR spectra, in terms of the time elapsed interaction with the protein. As “mimimum” for the definition of the interaction was used the distance of 4.0 Å, which is estimated as the limit for the detection of a significant signal from STD-NMR.

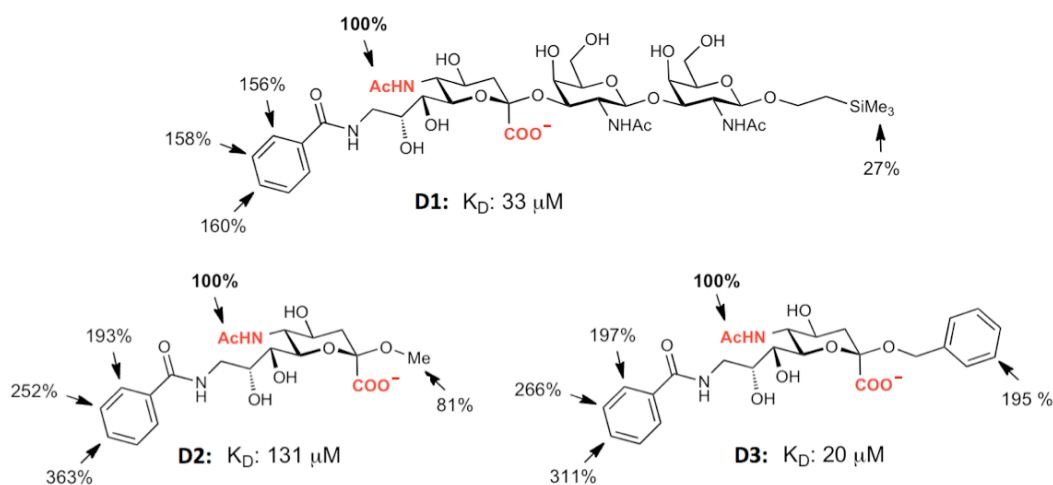


Figure 40: Structures of glycomimetics D1, D2, and D3. STD effects for the glycomimetics, D1, D2 and D3 in this study, as well as their K_D values determined using the *Biacore* technology, are indicated.

In order to minimize the bias from the initial docked position, as well as to compare the STD-NMR epitopes to the *in silico* modeling results, molecular dynamics (MD) simulations were performed for the glycomimetics D1, D2, and D3.

The comparison between the distances of the *N*-benzoate to the protein, and the data from STD-NMR shows a remarkable agreement. The frequency of interaction within the cut-off of 4.0 Å, sampled for 4 ns in MD simulations, reflects a higher contribution to the binding of the *N*-benzoate for ligands D2 and D3 (average of 83% and 86% respectively) when compared to ligand D1 (32%).

Moreover the percentages of minima for the averaged *ortho*, *meta* and *para* positions displayed a clear trend going from the lower interaction time of the

ortho position to the higher interaction time percentage of the *para* position, as detectable from the STD-NMR experiments. For compound D1, this difference is not as pronounced, both in the STD-NMR signal and in the percentage of estimated interaction time through MD simulation.

Table 25: Comparison of STD-NMR signals, frequency of interaction and average distances within the cut-off of 4.0 Å as derived from MD simulations, for compounds D1, D2, and D3. Ortho and meta positions are listed both as average and individually (ortho1/ortho2) referring to the two different faces of the *N*-benzoate.

Compound		STD-NMR signal (%)	MD % minima < 4.0 Å	Average distance (Å) minima < 4.0 Å
D1	ortho	156 %	26%	3.27
	ortho1/ortho2		27%/24%	
	meta	158 %	34 %	3.27
	meta1/meta2		36%/32%	
	para	160 %	36 %	3.22
D2	ortho	193 %	74 %	3.46
	ortho1/ortho2		81%/67%	
	meta	252 %	81 %	3.19
	meta1/meta2		88%/75%	
	para	363 %	94 %	3.44
D3	ortho	197 %	74 %	3.37
	ortho1/ortho2		76%/73%	
	meta	266 %	84 %	3.20
	meta1/meta2		85%/83%	
	para	311 %	98 %	3.37

In Figure 41 are shown the docking modes, from manual docking, for the glycomimetics D1 (A), D2 (B), and D3 (C). The stability of the C α carbons was monitored to evaluate the overall behavior of MAG when simulated in complex with the three compounds.

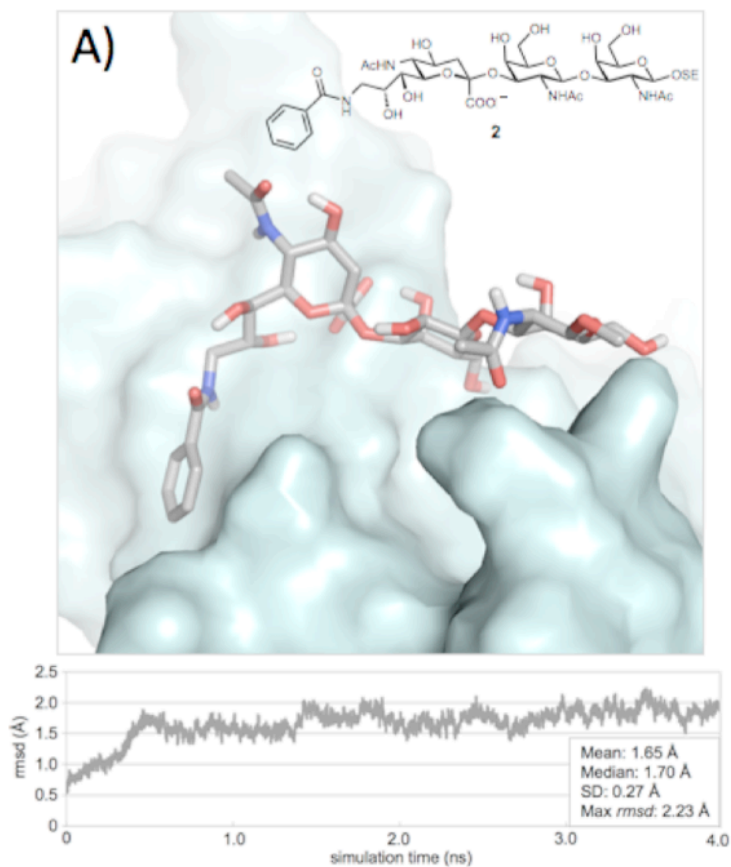


Figure 41: Detailed caption in the next page.

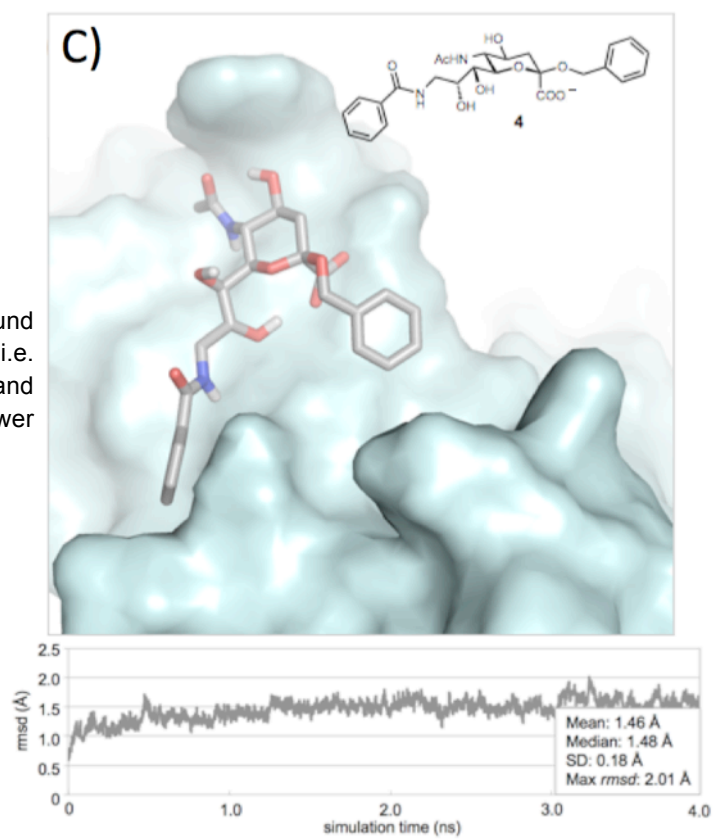
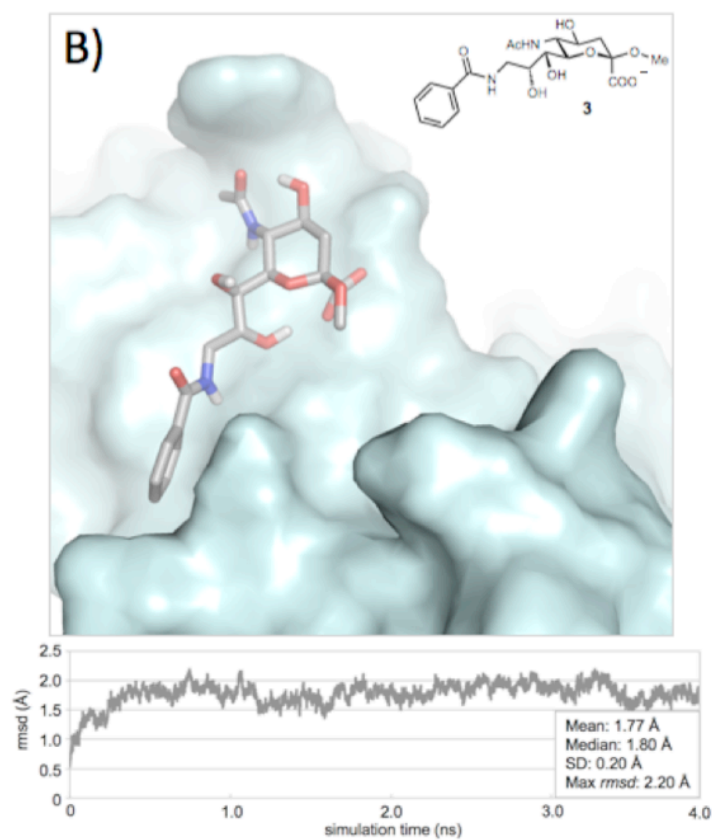


Figure 42:Devised binding modes for compound D1, D2 and D3. Relevant statistical data – i.e. mean, median, standard deviation (SD), and maximal reached *rmsd* – are listed in the lower right of each plot.

3.3 Estimation of relative-free energy differences using FEP

3.3.1 Series 3

Free energy perturbation calculations can be used to determine the relative free energy difference ($\Delta\Delta G$) between two equilibrium states.

FEP simulations of 1.0 ns were performed using the complex of MAG with one of the three compounds at the time as starting point (with one substituent on the indole ring), which was mutated to a similar ligand where the substituent was removed from the indole ring. During the FEP the ligand is simulated both in complex with the protein and in a solvent box (SPC water model). The estimation of the relative free energy difference between the two states ($\Delta\Delta G$) is obtained subtracting the ΔG calculated for the ligand interacting with the solvent (ΔG_2) to the ΔG calculated for the ligand in complex with the protein (ΔG_1).

The relative free energy difference for transitions from one state (with the substituent) to another one (without the substituent) is positive. Positive $\Delta\Delta G$ implies that the system before mutation has lower free energy than after mutation. FEP simulations suggest that A19/A20/A21 bind to MAG more strongly compared to A19/A20/A21 where the substituents were removed. The difference corresponds to a factor of 894. The absolute ΔG values can be difficultly compared since the calculations are not calibrated for this specific biological system, nonetheless the $\Delta\Delta G$ s for the three complexes are in a comparable range. Thus, the compounds with unsubstituted indole rings bind more weakly than the correspondent compounds with substituted indole rings, but the simulations are not able to discriminate the relative contribution of the three substituents.

Table 26: Relative free energy difference ($\Delta\Delta G$) for compound A19, A20, and A21 as calculated subtracting the complex with the protein (ΔG_1) and with the solvent (ΔG_2).

Compound	ΔG_1 _{complex} (kcal/mol)	ΔG_2 _{solvent} (kcal/mol)	$\Delta\Delta G$ (kcal/mol)
A-19	40.68±0.13	39.16±0.16	1.53±0.21
A-20	1.23±0.08	-1.26±0.09	2.49±0.12
A-21	-2.95±0.08	-5.01±0.08	2.07±0.12

Conclusions

1 Water in molecular simulations and drug design

A realistic simulation of ligands binding to proteins should account for structural and bridging solvent molecules. Based on a survey of 392 high-resolution crystal structures from the PDB, an average of 4.6 water molecules were identified at the ligand–protein interface. The 75% of these water molecules were found to be involved in polar interactions with both ligand and protein.

A rigorous analysis of small-molecule crystal structures from the CSD was performed to collect data on the geometry of interactions of water molecules when interacting with generic functional groups. *Ab initio* interaction energies between water and selected functional group representatives were calculated to construct an empiric hydration propensity scale. This information served the basis for an algorithm (AcquaAlta), which solvates ligand–protein interfaces. AcquaAlta is based on the directionality of hydrogen bonds and can solvate both classical hydrogen bond moieties as well as halogen and hydrophobic functionalities of the ligand.

The approach was validated on 20 X-ray structures with resolution ranging from 0.97 to 2.60 Å, checking the match of experimental and calculated water molecules. The match for binding site waters was 76%. The algorithm's accuracy was not influenced substantially by the resolution of the crystal structures. When we applied geometry and energy filters to identify only the water molecules having polar contacts with both ligand and protein (i.e. bridging waters), the match rose to 87.5%. WATGEN, an algorithm to model water networks at protein–protein surfaces, yielded comparable results predicting 72% and 88% of water sites placed within 1.5 and 2.0 Å, respectively.

Subsequently, the solvation approach was combined to a pharmacophore-based alignment tool and applied it to 14 structures of OppA. This approach yielded poses within a *rmsd* of 1.0 Å from the crystal structures for 13 out of 14 complexes. The match, without any filtering criteria, between experimental and calculated water molecules was 66% and 53% when based on the crystal and

docked poses, respectively. This drop in performance highlights the importance of a reasonable starting position to correctly reproduce water molecule positions. In a control test for the previously mentioned WATGEN algorithm, solvent molecules randomly placed at protein–protein interfaces were able to match only 22% and 40% of the experimental water sites within 1.5 and 2.0 Å.

In *AcquaAlta*, the solvated ligand-protein complex is not evaluated based on energy or on modification of entropy and enthalpy of the binding, but rather on evaluation of the presence of hydrogen-bond partners that a water molecule could bridge and thus bring to favorable interaction. The aim of this solvation algorithm is to produce crystal-like binding poses with optimally arranged bridging waters, expecting that further refinement using force-field minimization routines would benefit and lead to a more accurate evaluation of thermodynamics of ligand binding. *AcquaAlta* is thought to be applied to docking studies where the presence or absence of a water molecule, calculated on each ligand conformation, can substantially affect pose scoring. On one side, this could help obtaining reasonable binding modes; on the other one, the geometric criteria used do not give any information whether the water enhances or worsens the binding.

2 MeTOX: Metabolism and TOXicology prediction

The aim of this part of the thesis was to establish a reliable approach for the prediction of drug-like molecules binding to CYP2C9 and CYP2D6. The promiscuity of these enzymes is often attributed to their flexibility and their large active-site volume. Comparison of the mobility of C α carbons during the X-ray determination and during MD simulations confirmed the high flexibility of the external loops. On the contrary the C α carbons of the active site are relatively rigid. In order to correctly simulate the binding to CYP2D6 and CYP2C9, side-chain flexibility was accounted during the pre-alignment and docking protocol. The alignment of a 4D ensemble of 56 compounds for CYP2D6 and 85 for CYP2C9 was used to generate mQSAR models (software Quasar). These models represent the active site as a surrogate characterized by quasi-atomistic properties reflecting those of the true biological receptor.

The models converged at a cross-validated r^2 of 0.811 for CYP2D6 and 0.695 for CYP2C9. The lower value of predictive r^2 for CYP2C9 (*i.e.* 0.423) compared to CYP2D6 (*i.e.* 0.711) is due to the high concentration of compounds displaying low binding affinities as well as to the narrow range of affinities. Nonetheless, the worst predicted compound in the cross-validation procedure is within a factor of 5.0 for CYP2D6 and 4.4 for CYP2C9. Model sensitivity was assessed by 40 scramble tests for each of the two isoforms, which yielded an average predictive r^2 of -0.318 for CYP2D6, and -0.283 for CYP2C9. The visual inspection of the docked poses confirmed consistency between the location of sites of metabolism obtained from experimental studies and the linear accessibility of these sites to the heme in the docked complexes. When models were tested on an external dataset, the experimental values were predicted within a factor of 10 for the 60% of the compounds for CYP2D6 and for the 80% for CYP2C9 (86.7% and 87.5% within a factor of 25 for CYP2D6 and CYP2C9, respectively). The binding properties for 64 either non-binders or weak-binders for CYP2D6 and 56 for CYP2C9, allowed us to test the probability to obtain false positive alerts from our models. In this account, 85.9% and 82.1% were predicted to be non-binders or within a factor of 10 from the weakest binding compound from the training set.

The two CYPs models have been added to the *VirtualToxLab* developed by the Biographics Laboratory 3R for the *in silico* prediction of metabolic and endocrine disruption and interference with the hERG ion channel. Presently, it includes 16 validated models for 5 CYPs, 10 nuclear receptors and the hERG ion channel.

This study underlines that for the correct modeling of drug-like molecules binding to CYP2D6 and CYP2C9, ligand–pharmacophore features, oxidation state of the iron cofactor, and active-site motion must be considered. Previously published models are either based on pharmacophore hypotheses or physico-chemical and structural descriptors. Since the binding affinity of small-molecule ligands depends highly on the protein–ligand interaction and the experimental structures of both CYP2D6 and CYP2C9 are available, it is advantageous to identify the binding mode through flexible docking at these entities. This allows to simulate local induced fit. The mQSAR technology used for estimating binding affinities would seem to be necessary as it allows to quantify protein–ligand interaction at the 4D structural level but also induced fit (5D) and solvation phenomena (6D). Our results suggest that mixed-model QSAR including ligand pre-alignment is a reliable tool for predicting sites of metabolism and particularly for estimating associated binding affinities but also for producing toxicity alerts.

3 Computer-aided drug design at the Myelin Associated Glycoprotein (MAG)

The inability of the CNS in its regeneration is connected to the presence of different inhibitors protein such as the Myelin Associated Glycoprotein (MAG) and for the presence of astrocytes, which are recruited on the site of injury.

The design of MAG antagonists represents a valuable tool for the investigation of such a protein behavior and its potential to enhance spinal cord regeneration. Previous studies reported that structurally related gangliosides (GD1a, GT1b, and GQ1b α) are functional ligands for MAG.¹²⁷ In this thesis different glycomimetic structures were investigated through molecular docking, molecular dynamics and free-energy perturbation simulations.

Unfortunately, no 3D structure of MAG is yet available. Therefore a homology model of MAG was generated and used to perform the above listed simulations.

In terms of docking modes all the compounds were docked aiming to establish a precise hydrogen bond network with residues, which were demonstrated to be relevant for binding. These residues are primarily Arg118, Gln126 and Thr128.

Molecular dynamics simulations were used to produce a simulation of the kinetics of binding. Moreover the homology model of MAG is possibly not presenting the exact orientations of protein side-chains, as they would be found in its physiological environment. Therefore, MD can explore more side-chains configurations with a higher possibility to simulate a correct configuration when MAG is in complex with novel inhibitors.

Free-energy perturbation simulations were used to evaluate relative-free energy differences between two ligands. Ligands were mutated in order to measure the relative importance of single functional groups to the binding.

Being aware of the weaknesses and limitations of applying molecular modeling techniques on structures obtained through homology modeling, all the simulations were aiming to consider this uncertainty of the starting atomic coordinates of the protein and to give a possible explanation of the experimental data obtained. An X-ray structure of MAG would be beneficial to evaluate the quality of these predictions and to further proceed on structure-based design of potent inhibitors.

References

1. van de Waterbeemd, H.; Gifford, E., ADMET in silico modelling: towards prediction paradise? *Nat Rev Drug Discov* **2003**, 2, (3), 192-204.
2. Hubbard, S. J.; Gross, K. H.; Argos, P., Intramolecular cavities in globular proteins. *Protein Eng* **1994**, 7, (5), 613-26.
3. Matthews, B. W.; Liu, L., A review about nothing: are apolar cavities in proteins really empty? *Protein Sci* **2009**, 18, (3), 494-502.
4. Quioco, F.; Wilson, D.; Vyas, N., Substrate specificity and affinity of a protein modulated by bound water molecules. *Nature* **1989**, 340, (6232), 404-7.
5. Babor, M.; Sobolev, V.; Edelman, M., Conserved positions for ribose recognition: importance of water bridging interactions among ATP, ADP and FAD-protein complexes. *J Mol Biol* **2002**, 323, (3), 523-32.
6. Mikol, V.; Kallen, J.; Pflügl, G.; Walkinshaw, M. D., X-ray structure of a monomeric cyclophilin A-cyclosporin A crystal complex at 2.1 Å resolution. *J Mol Biol* **1993**, 234, (4), 1119-30.
7. Zheng, J.; Trafny, E. A.; Knighton, D. R.; Xuong, N. H.; Taylor, S. S.; Ten Eyck, L. F.; Sowadski, J. M., 2.2 Å refined crystal structure of the catalytic subunit of cAMP-dependent protein kinase complexed with MnATP and a peptide inhibitor. *Acta Crystallogr D Biol Crystallogr* **1993**, 49, (Pt 3), 362-5.
8. Bhat, T. N.; Bentley, G. A.; Boulot, G.; Greene, M. I.; Tello, D.; Dall'Acqua, W.; Souchon, H.; Schwarz, F. P.; Mariuzza, R. A.; Poljak, R. J., Bound water molecules and conformational stabilization help mediate an antigen-antibody association. *Proc Natl Acad Sci U S A* **1994**, 91, (3), 1089-93.
9. Joachimiak, A.; Haran, T. E.; Sigler, P. B., Mutagenesis supports water mediated recognition in the trp repressor-operator system. *EMBO J* **1994**, 13, (2), 367-72.
10. Huang, K.; Lu, W.; Anderson, S.; Laskowski, M.; James, M. N., Water molecules participate in proteinase-inhibitor interactions: crystal structures of Leu18, Ala18, and Gly18 variants of turkey ovomucoid inhibitor third domain complexed with *Streptomyces griseus* proteinase B. *Protein Sci* **1995**, 4, (10), 1985-97.
11. Holdgate, G. A.; Tunncliffe, A.; Ward, W. H.; Weston, S. A.; Rosenbrock, G.; Barth, P. T.; Taylor, I. W.; Pauptit, R. A.; Timms, D., The entropic penalty of ordered water accounts for weaker binding of the antibiotic novobiocin to a resistant mutant of DNA gyrase: a thermodynamic and crystallographic study. *Biochemistry* **1997**, 36, (32), 9663-73.
12. <http://www.pdb.org>
13. Lu, Y.; Wang, R.; Yang, C.; Wang, S., Analysis of ligand-bound water molecules in high-resolution crystal structures of protein-ligand complexes. *J Chem Inf Model* **2007**, 47, (2), 668-75.

14. Li, Z.; Lazaridis, T., Water at biomolecular binding interfaces. *Phys. Chem. Chem. Phys.* **2007**, *9*, 573-581.
15. Michel, J.; Tirado-Rives, J.; Jorgensen, W., Prediction of the water content in protein binding sites. *J Phys Chem B* **2009**, *113*, (40), 13337-46.
16. Ajay; Murcko, M. A., Computational methods to predict binding free energy in ligand-receptor complexes. *J Med Chem* **1995**, *38*, (26), 4953-67.
17. Leach, A. R.; Shoichet, B. K.; Peishoff, C. E., Prediction of protein-ligand interactions. Docking and scoring: successes and gaps. *J Med Chem* **2006**, *49*, (20), 5851-5.
18. Mancera, R. L., Molecular modeling of hydration in drug design. *Curr Opin Drug Discov Devel* **2007**, *10*, (3), 275-80.
19. Ladbury, J., Just add water! The effect of water on the specificity of protein-ligand binding sites and its potential application to drug design. *Chem Biol* **1996**, *3*, (12), 973-80.
20. Dunitz, J. D., The entropic cost of bound water in crystals and biomolecules. *Science* **1994**, *264*, (5159), 670.
21. Hamelberg, D.; McCammon, J. A., Standard free energy of releasing a localized water molecule from the binding pockets of proteins: double-decoupling method. *J Am Chem Soc* **2004**, *126*, (24), 7683-9.
22. Poornima, C.; Dean, P., Hydration in drug design. 1. Multiple hydrogen-bonding features of water molecules in mediating protein-ligand interactions. *J Comput Aided Mol Des* **1995**, *9*, (6), 500-12.
23. Poornima, C.; Dean, P., Hydration in drug design. 2. Influence of local site surface shape on water binding. *J Comput Aided Mol Des* **1995**, *9*, (6), 513-20.
24. Poornima, C.; Dean, P., Hydration in drug design. 3. Conserved water molecules at the ligand-binding sites of homologous proteins. *J Comput Aided Mol Des* **1995**, *9*, (6), 521-31.
25. Kortvelyesi, T.; Dennis, S.; Silberstein, M.; Brown, L. r.; Vajda, S., Algorithms for computational solvent mapping of proteins. *Proteins* **2003**, *51*, (3), 340-51.
26. Ehrlich, L.; Reczko, M.; Bohr, H.; Wade, R., Prediction of protein hydration sites from sequence by modular neural networks. *Protein Eng* **1998**, *11*, (1), 11-9.
27. Lam, P.; Jadhav, P.; Eyermann, C.; Hodge, C.; Ru, Y.; Bacheler, L.; Meek, J.; Otto, M.; Rayner, M.; Wong, Y., Rational design of potent, bioavailable, nonpeptide cyclic ureas as HIV protease inhibitors. *Science* **1994**, *263*, (5145), 380-4.
28. Watson, K. A.; Mitchell, E. P.; Johnson, L. N.; Son, J. C.; Bichard, C. J.; Orchard, M. G.; Fleet, G. W.; Oikonomakos, N. G.; Leonidas, D. D.; Kontou, M., Design of inhibitors of glycogen phosphorylase: a study of alpha- and beta-C-glucosides and 1-thio-beta-D-glucose compounds. *Biochemistry* **1994**, *33*, (19), 5745-58.

29. Chen, J. M.; Xu, S. L.; Wawrzak, Z.; Basarab, G. S.; Jordan, D. B., Structure-based design of potent inhibitors of scytalone dehydratase: displacement of a water molecule from the active site. *Biochemistry* **1998**, 37, (51), 17735-44.
30. Liu, C.; Wroblewski, S. T.; Lin, J.; Ahmed, G.; Metzger, A.; Wityak, J.; Gillooly, K. M.; Shuster, D. J.; McIntyre, K. W.; Pitt, S.; Shen, D. R.; Zhang, R. F.; Zhang, H.; Doweyko, A. M.; Diller, D.; Henderson, I.; Barrish, J. C.; Dodd, J. H.; Schieven, G. L.; Leftheris, K., 5-Cyanopyrimidine derivatives as a novel class of potent, selective, and orally active inhibitors of p38alpha MAP kinase. *J Med Chem* **2005**, 48, (20), 6261-70.
31. Seo, J.; Igarashi, J.; Li, H.; Martasek, P.; Roman, L. J.; Poulos, T. L.; Silverman, R. B., Structure-based design and synthesis of N(omega)-nitro-L-arginine-containing peptidomimetics as selective inhibitors of neuronal nitric oxide synthase. Displacement of the heme structural water. *J Med Chem* **2007**, 50, (9), 2089-99.
32. Campiani, G.; Kozikowski, A. P.; Wang, S.; Ming, L.; Nacci, V.; Saxena, A.; Doctor, B. P., Synthesis and anticholinesterase activity of huperzine A analogues containing phenol and catechol replacements for the pyridone ring. *Bioorg Med Chem Lett* **1998**, 8, (11), 1413-8.
33. Goldbaum, F. A.; Schwarz, F. P.; Eisenstein, E.; Cauherhff, A.; Mariuzza, R. A.; Poljak, R. J., The effect of water activity on the association constant and the enthalpy of reaction between lysozyme and the specific antibodies D1.3 and D44.1. *J Mol Recognit* **1996**, 9, (1), 6-12.
34. Tame, J. R.; Sleigh, S. H.; Wilkinson, A. J.; Ladbury, J. E., The role of water in sequence-independent ligand binding by an oligopeptide transporter protein. *Nat Struct Biol* **1996**, 3, (12), 998-1001.
35. Guillot, B., A reappraisal of what we have learnt during three decades of computer simulations on water. *Journal of Molecular Liquids* **2002**, 101, (1-3), 219-260.
36. <http://www.lsbu.ac.uk/water/>
37. Berendsen, H.; Postma, J.; Van Gunsteren, W.; Hermans, J., Interaction Models for Water in Relation to Protein Hydration. *Intermolecular Forces* **1981**, 331-342.
38. Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P., The missing term in effective pair potentials. *J. Phys. Chem.* **1987**, 91, (24), 6269-6271.
39. Jorgensen, W. L.; Madura, J. D., Solvation and Conformation of Methanol in Water. *JACS* **1983**, 105, (6), 1407-1413.
40. Mahoney, M. W.; Jorgensen, W. L., **A Five-Site Model for Liquid Water and the Reproduction of the Density Anomaly by Rigid, Nonpolarizable Potential Functions.** *J. Chem. Phys.* **2000**, 112, 8910-8922.
41. Caleman, C. http://xray.bmc.uu.se/~calle/md_phd/water_models.pdf
42. Levitt, M.; Park, B. H., Water: now you see it, now you don't. *Structure* **1993**, 1, (4), 223-6.

43. Meyer, E., Internal water molecules and H-bonding in biological macromolecules: a review of structural features with functional implications. *Protein Sci* **1992**, 1, (12), 1543-62.
44. Murray-Rust, P.; Glusker, J. P., Directional hydrogen bonding to sp²- and sp³-hybridized oxygen atoms and its relevance to ligand-macromolecule interactions. *J. Am. Chem. Soc.*, **1984**, 106, (4), 1018-1025.
45. Taylor, R.; Kennard, O.; Versichel, W., Geometry of the imino-carbonyl (N-H...O:C) hydrogen bond. 1. Lone-pair directionality. *J. Am. Chem. Soc.* **1983**, 105, (18), 5761-5766.
46. Vedani, A.; Dunitz, J. D., Lone-pair directionality in hydrogen-bond potential functions for molecular mechanics calculations: the inhibition of human carbonic anhydrase II by sulfonamides. *J. Am. Chem. Soc.* **1985**, 107, (25), 7653-7658.
47. Lommerse, J. P. M.; Price, S. L.; Taylor, R., Hydrogen bonding of carbonyl, ether, and ester oxygen atoms with alkanol hydroxyl groups. *J. Comput. Chem.* **1997**, 18, 757-774.
48. Nobeli, I.; Price, S. L.; Lommerse, J. P. M.; Taylor, R., Hydrogen bonding properties of oxygen and nitrogen acceptors in aromatic heterocycles. *J. Comput. Chem.* **1997**, 18, 2060-2074.
49. Bissantz, C.; Kuhn, B.; Stahl, M., A Medicinal Chemist's Guide to Molecular Interactions. *J Med Chem* **2010**.
50. Ho, B. K.; Gruswitz, F., HOLLOW: generating accurate representations of channel and interior surfaces in molecular structures. *BMC Struct Biol* **2008**, 8, 49.
51. Pymol. <http://www.pymol.org>
52. Denisov, V. P.; Venu, K.; Peters, J.; Hörlein, H. D.; Halle, B., Orientational Disorder and Entropy of Water in Protein Cavities. *J. Phys. Chem. B.* **1997**, 101, (45), 9380-9389.
53. Henn, A. R.; Kauzmann, W., Equation of state of a random network, continuum model of liquid water. *J. Phys. Chem.* **1989**, 93, (9), 3770-3783.
54. Chaplin, M. F., A proposal for the structuring of water. *Biophys Chem* **2000**, 83, (3), 211-21.
55. <http://www.lsbu.ac.uk/water/anmlies.html>
56. Amadasi, A.; Spyraakis, F.; Cozzini, P.; Abraham, D.; Kellogg, G.; Mozzarelli, A., Mapping the energetics of water-protein and water-ligand interactions with the "natural" HINT forcefield: predictive tools for characterizing the roles of water in biomolecules. *J Mol Biol* **2006**, 358, (1), 289-309.
57. García-Sosa, A.; Mancera, R.; Dean, P., WaterScore: a novel method for distinguishing between bound and displaceable water molecules in the crystal structure of the binding site of protein-ligand complexes. *J Mol Model* **2003**, 9, (3), 172-82.
58. Verdonk, M.; Chessari, G.; Cole, J.; Hartshorn, M.; Murray, C.; Nissink, J.; Taylor, R.; Taylor, R., Modeling water molecules in protein-ligand docking using GOLD. *J Med Chem* **2005**, 48, (20), 6504-15.

59. Friesner, R.; Murphy, R.; Repasky, M.; Frye, L.; Greenwood, J.; Halgren, T.; Sanschagrin, P.; Mainz, D., Extra precision glide: docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. *J Med Chem* **2006**, 49, (21), 6177-96.
60. Corbeil, C.; Englebienne, P.; Moitessier, N., Docking ligands into flexible and solvated macromolecules. 1. Development and validation of FITTED 1.0. *J Chem Inf Model* **2007**, 47, (2), 435-49.
61. Osterberg, F.; Morris, G.; Sanner, M.; Olson, A.; Goodsell, D., Automated docking to multiple target structures: incorporation of protein mobility and structural water heterogeneity in AutoDock. *Proteins* **2002**, 46, (1), 34-40.
62. Rejto, P. A.; Verkhivker, G. M., Mean field analysis of FKBP12 complexes with FK506 and rapamycin: implications for a role of crystallographic water molecules in molecular recognition and specificity. *Proteins* **1997**, 28, (3), 313-24.
63. Huang, N.; Shoichet, B., Exploiting ordered waters in molecular docking. *J Med Chem* **2008**, 51, (16), 4862-5.
64. Abel, R.; Young, T.; Farid, R.; Berne, B.; Friesner, R., Role of the active-site solvent in the thermodynamics of factor Xa ligand binding. *J Am Chem Soc* **2008**, 130, (9), 2817-31.
65. Thilagavathi, R.; Mancera, R., Ligand-protein cross-docking with water molecules. *J Chem Inf Model* **2010**, 50, (3), 415-21.
66. Birch, L.; Murray, C.; Hartshorn, M.; Tickle, I.; Verdonk, M., Sensitivity of molecular docking to induced fit effects in influenza virus neuraminidase. *J Comput Aided Mol Des* **2002**, 16, (12), 855-69.
67. Denisov, I. G.; Makris, T. M.; Sligar, S. G.; Schlichting, I., Structure and Chemistry of Cytochrome P450. *Chem. Rev.* **2005**, 105, 2253-2277.
68. Omura, T., Forty years of cytochrome P450. *Biochem Biophys Res Commun* **1999**, 266, (3), 690-8.
69. Guengerich, F. P., Common and uncommon cytochrome P450 reactions related to metabolism and chemical toxicity. *Chem Res Toxicol* **2001**, 14, (6), 611-50.
70. Nelson, D. R.; Koymans, L.; Kamataki, T.; Stegeman, J. J.; Feyereisen, R.; Waxman, D. J.; Waterman, M. R.; Gotoh, O.; Coon, M. J.; Estabrook, R. W.; Gunsalus, I. C.; Nebert, D. W., P450 superfamily: update on new sequences, gene mapping, accession numbers and nomenclature. *Pharmacogenetics* **1996**, 6, (1), 1-42.
71. Smith, D. A.; Abel, S. M.; Hyland, R.; Jones, B. C., Human cytochrome P450s: selectivity and measurement in vivo. *Xenobiotica* **1998**, 28, (12), 1095-128.
72. Shimada, T.; Yamazaki, H.; Mimura, M.; Inui, Y.; Guengerich, F. P., Interindividual variations in human liver cytochrome P-450 enzymes involved in the oxidation of drugs, carcinogens and toxic chemicals: studies with liver microsomes of 30 Japanese and 30 Caucasians. *J Pharmacol Exp Ther* **1994**, 270, (1), 414-23.

73. Rendic, S., Summary of information on human CYP enzymes: human P450 metabolism data. *Drug Metab Rev* **2002**, 34, (1-2), 83-448.
74. Miners, J. O.; Birkett, D. J., Cytochrome P450C9: an enzyme of major importance in human drug metabolism. *Br J Clin Pharmacol* **1998**, 45, (6), 525-38.
75. Jones, C. R.; Guengerich, F. P.; Rice, J. M.; Lubet, R. A., Induction of various cytochromes CYP2B, CYP2C and CYP3A by phenobarbitone in non-human primates. *Pharmacogenetics* **1992**, 2, (4), 160-72.
76. J., S. M.; Breckinridge, A. M., Drug interactions with warfarin. *Drugs* **1983**, 25, 610-620.
77. O'Reilly, R. A.; Goulart, D. A.; Kunze, K. L.; Neal, J.; Gibaldi, M.; Eddy, A. C.; Trager, W. F., Mechanisms of the stereoselective interaction between miconazole and racemic warfarin in human subjects. *Clin Pharmacol Ther* **1992**, 51, (6), 656-67.
78. Transon, C.; Leemann, T.; Dayer, P., In vitro comparative inhibition profiles of major human drug metabolising cytochrome P450 isozymes (CYP2C9, CYP2D6 and CYP3A4) by HMG-CoA reductase inhibitors. *Eur J Clin Pharmacol* **1996**, 50, (3), 209-15.
79. Zhou, S. F.; Liu, J. P.; Chowbay, B., **Polymorphism of human cytochrome P450 enzymes and its clinical impact.** *Drug. Metab. Rev* **2009**, 41, (2), 89-295.
80. de Groot, M. J.; Ekins, S., Pharmacophore modeling of cytochromes P450. *Adv Drug Deliv Rev* **2002**, 54, (3), 367-83.
81. Korhonen, L. E.; Rahnasto, M.; Mähönen, N. J.; Wittekindt, C.; Poso, A.; Juvonen, R. O.; Raunio, H., Predictive three-dimensional quantitative structure-activity relationship of cytochrome P450 1A2 inhibitors. *J Med Chem* **2005**, 48, (11), 3808-15.
82. Ekins, S.; Bravi, G.; Wikel, J. H.; Wrighton, S. A., Three-dimensional-quantitative structure activity relationship analysis of cytochrome P-450 3A4 substrates. *J Pharmacol Exp Ther* **1999**, 291, (1), 424-33.
83. Ekins, S.; Bravi, G.; Ring, B. J.; Gillespie, T. A.; Gillespie, J. S.; Vandenbranden, M.; Wrighton, S. A.; Wikel, J. H., Three-dimensional quantitative structure activity relationship analyses of substrates for CYP2B6. *J Pharmacol Exp Ther* **1999**, 288, (1), 21-9.
84. Ekins, S.; De Groot, M. J.; Jones, J. P., Pharmacophore and three-dimensional quantitative structure activity relationship methods for modeling cytochrome P450 active sites. *Drug Metabolism and Disposition* **2001**, 29, (7), 936-944.
85. Ekins, S.; Bravi, G.; Binkley, S.; Gillespie, J. S.; Ring, B. J.; Wikel, J. H.; Wrighton, S. A., Three- and four-dimensional-quantitative structure activity relationship (3D/4D-QSAR) analyses of CYP2C9 inhibitors. *Drug Metab Dispos* **2000**, 28, (8), 994-1002.
86. Ridderström, M.; Zamora, I.; Fjellström, O.; Andersson, T. B., Analysis of selective regions in the active sites of human cytochromes P450, 2C8, 2C9,

- 2C18, and 2C19 homology models using GRID/CPCA. *J Med Chem* **2001**, 44, (24), 4072-81.
87. Lewis, D. F.; Sams, C.; Loizou, G. D., A quantitative structure-activity relationship analysis on a series of alkyl benzenes metabolized by human cytochrome p450 2E1. *J Biochem Mol Toxicol* **2003**, 17, (1), 47-52.
88. Hritz, J.; de Ruiter, A.; Oostenbrink, C., Impact of plasticity and flexibility on docking results for cytochrome P450 2D6: a combined approach of molecular dynamics and ligand docking. *J Med Chem* **2008**, 51, (23), 7469-77.
89. Williams, P. A.; Cosme, J.; Vinkovic, D. M.; Ward, A.; Angove, H. C.; Day, P. J.; Vonrhein, C.; Tickle, I. J.; Jhoti, H., Crystal structures of human cytochrome P450 3A4 bound to metyrapone and progesterone. *Science* **2004**, 305, (5684), 683-6.
90. Williams, P. A.; Cosme, J.; Ward, A.; Angova, H. C.; Vinkovic, D. M.; Jhoti, H., Crystal structure of human cytochrome P4502C9 with bound warfarin. *Nature* **2003**, 424, (6947), 464-468.
91. Guengerich, F. P., A malleable catalyst dominates the metabolism of drugs. *Proceedings of the National Academy of Sciences of the United States of America* **2006**, 103, (37), 13565-13566.
92. Ekroos, M.; Sjögren, T., Structural basis for ligand promiscuity in cytochrome P450 3A4. *Proc Natl Acad Sci U S A* **2006**, 103, (37), 13682-7.
93. Scott, E. E.; He, Y. A.; Wester, M. R.; White, M. A.; Chin, C. C.; Halpert, J. R.; Johnson, E. F.; Stout, C. D., An open conformation of mammalian cytochrome P450 2B4 at 1.6-Å resolution. *Proc Natl Acad Sci U S A* **2003**, 100, (23), 13196-201.
94. Scott, E. E.; White, M. A.; He, Y. A.; Johnson, E. F.; Stout, C. D.; Halpert, J. R., Structure of mammalian cytochrome P450 2B4 complexed with 4-(4-chlorophenyl)imidazole at 1.9-Å resolution: insight into the range of P450 conformations and the coordination of redox partner binding. *J Biol Chem* **2004**, 279, (26), 27294-301.
95. Zhao, Y.; White, M. A.; Muralidhara, B. K.; Sun, L.; Halpert, J. R.; Stout, C. D., Structure of microsomal cytochrome P450 2B4 complexed with the antifungal drug bifonazole: insight into P450 conformational plasticity and membrane interaction. *J Biol Chem* **2006**, 281, (9), 5973-81.
96. Leach, A., *Molecular Modelling. Principles and applications*. 2nd ed.; Pearson Education Limited: Essex, England, 2001.
97. Cramer, I. R.; Patterson, D.; Bunce, J., Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. *J. Am. Chem. Soc.* **1988**, 110, 5959-5967.
98. Rahnasto, M.; Raunio, H.; Poso, A.; Wittekindt, C.; Juvonen, R. O., Quantitative structure-activity relationship analysis of inhibitors of the nicotine metabolizing CYP2A6 enzyme. *J Med Chem* **2005**, 48, (2), 440-9.
99. Chohan, K.; Paine, S.; Waters, N., Quantitative structure activity relationships in drug metabolism. *Curr Top Med Chem* **2006**, 6, 1569-78.

100. Dearden, J.; Cronin, M.; Kaiser, K., How not to develop a quantitative structure-activity or structure-property relationship (QSAR/QSAPR). *SAR and QSAR in Environmental Research* **2009**, 20, 241–266.
101. Golbraikh, A.; Tropsha, A., Beware of q(2)! *Journal of Molecular Graphics & Modelling* **2002**, 20, 269-276.
102. Klebe, G.; Abraham, U.; Mietzner, T., Molecular similarity indices in a comparative analysis (CoMSIA) of drug molecules to correlate and predict their biological activity. *J Med Chem* **1994**, 37, (24), 4130-46.
103. Kubinyi, H.; Hamprecht, F. A.; Mietzner, T., Three-dimensional quantitative similarity-activity relationships (3D QSiAR) from SEAL similarity matrices. *J Med Chem* **1998**, 41, (14), 2553-64.
104. Höltje, H.; Sippl, W.; Rognan, D.; Folkers, G., *Molecular modeling. Basic principles and applications*. Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany, 2003.
105. McMasters, D. R.; Torres, R. A.; Crathern, S. J.; Dooney, D. L.; Nachbar, R. B.; Sheridan, R. P.; Korzekwa, K. R., Inhibition of recombinant cytochrome p450 isoforms 2D6 and 2C9 by diverse drug-like molecules. *Journal of Medicinal Chemistry* **2007**, 50, (14), 3205-3213.
106. Vedani, A.; Dobler, M., Multidimensional QSAR: Moving from three- to five-dimensional concepts. *Quantitative Structure-Activity Relationships* **2002**, 21, (4), 382-390.
107. Vedani, A.; Dobler, M., 5D-QSAR: the key for simulating induced fit? *J Med Chem* **2002**, 45, (11), 2139-49.
108. Vedani, A.; Dobler, M., Multi-dimensional QSAR in drug research - Predicting binding affinities, toxicity and pharmacokinetic parameters. *Progress in Drug Research, Vol 55* **2000**, 55, 105-135.
109. Vedani, A.; Smiesko, M.; Spreafico, M.; Peristera, O.; Dobler, M., VirtualToxLab - in silico prediction of the toxic (endocrine-disrupting) potential of drugs, chemicals and natural products. Two years and 2,000 compounds of experience: a progress report. *Altex* **2009**, 26, (3), 167-76.
110. Qiu, J.; Cai, D.; Filbin, M. T., Glial inhibition of nerve regeneration in the mature mammalian CNS. *Glia* **2000**, 29, (2), 166-74.
111. Schwab, M. E.; Bartholdi, D., Degeneration and regeneration of axons in the lesioned spinal cord. *Physiol Rev* **1996**, 76, (2), 319-70.
112. Spencer, T.; Domeniconi, M.; Cao, Z.; Filbin, M. T., New roles for old proteins in adult CNS axonal regeneration. *Curr Opin Neurobiol* **2003**, 13, (1), 133-9.
113. Scherer, S. S.; Salzer, J. L., *Glial Cell Development; Basic Principles and Clinical Relevance*. Bios Scientific Publishers: Oxford, 1996.
114. Filbin, M. T., Myelin-associated inhibitors of axonal regeneration in the adult mammalian CNS. *Nat Rev Neurosci* **2003**, 4, (9), 703-13.
115. Mukhopadhyay, G.; Doherty, P.; Walsh, F. S.; Crocker, P. R.; Filbin, M. T., A novel role for myelin-associated glycoprotein as an inhibitor of axonal regeneration. *Neuron* **1994**, 13, (3), 757-67.

116. Cai, D.; Qiu, J.; Cao, Z.; McAtee, M.; Bregman, B. S.; Filbin, M. T., Neuronal cyclic AMP controls the developmental loss in ability of axons to regenerate. *J Neurosci* **2001**, 21, (13), 4731-9.
117. Quarles, R. H., Myelin-associated glycoprotein (MAG): past, present and beyond. *J Neurochem* **2007**, 100, (6), 1431-48.
118. Kelm, S.; Pelz, A.; Schauer, R.; Filbin, M. T.; Tang, S.; de Bellard, M. E.; Schnaar, R. L.; Mahoney, J. A.; Hartnell, A.; Bradfield, P., Sialoadhesin, myelin-associated glycoprotein and CD22 define a new family of sialic acid-dependent adhesion molecules of the immunoglobulin superfamily. *Curr Biol* **1994**, 4, (11), 965-72.
119. Yang, L. J.; Zeller, C. B.; Shaper, N. L.; Kiso, M.; Hasegawa, A.; Shapiro, R. E.; Schnaar, R. L., Gangliosides are neuronal ligands for myelin-associated glycoprotein. *Proc Natl Acad Sci U S A* **1996**, 93, (2), 814-8.
120. Tang, S.; Shen, Y. J.; DeBellard, M. E.; Mukhopadhyay, G.; Salzer, J. L.; Crocker, P. R.; Filbin, M. T., Myelin-associated glycoprotein interacts with neurons via a sialic acid binding site at ARG118 and a distinct neurite inhibition site. *J Cell Biol* **1997**, 138, (6), 1355-66.
121. Vinson, M.; Strijbos, P. J.; Rowles, A.; Facci, L.; Moore, S. E.; Simmons, D. L.; Walsh, F. S., Myelin-associated glycoprotein interacts with ganglioside GT1b. A mechanism for neurite outgrowth inhibition. *J Biol Chem* **2001**, 276, (23), 20280-5.
122. Wörter, V.; Schweigreiter, R.; Kinzel, B.; Mueller, M.; Barske, C.; Böck, G.; Frenzel, S.; Bandtlow, C. E., Inhibitory activity of myelin-associated glycoprotein on sensory neurons is largely independent of NgR1 and NgR2 and resides within Ig-Like domains 4 and 5. *PLoS One* **2009**, 4, (4), e5218.
123. Laurén, J.; Hu, F.; Chin, J.; Liao, J.; Airaksinen, M. S.; Strittmatter, S. M., Characterization of myelin ligand complexes with neuronal Nogo-66 receptor family members. *J Biol Chem* **2007**, 282, (8), 5715-25.
124. Robak, L. A.; Venkatesh, K.; Lee, H.; Raiker, S. J.; Duan, Y.; Lee-Osbourne, J.; Hofer, T.; Mage, R. G.; Rader, C.; Giger, R. J., Molecular basis of the interactions of the Nogo-66 receptor and its homolog NgR2 with myelin-associated glycoprotein: development of NgROMNI-Fc, a novel antagonist of CNS myelin inhibition. *J Neurosci* **2009**, 29, (18), 5768-83.
125. Quarles, R. H., A hypothesis about the relationship of myelin-associated glycoprotein's function in myelinated axons to its capacity to inhibit neurite outgrowth. *Neurochem Res* **2009**, 34, (1), 79-86.
126. Ernst, B.; Schwardt, O.; Mesch, S.; Wittwer, M.; Rossato, G.; Vedani, A., From the ganglioside GQ1b α to glycomimetic antagonists of the myelin-associated glycoprotein (MAG). *Chimia (Aarau)* **2010**, 64, (1-2), 17-22.
127. Collins, B. E.; Yang, L. J.; Mukhopadhyay, G.; Filbin, M. T.; Kiso, M.; Hasegawa, A.; Schnaar, R. L., Sialic acid specificity of myelin-associated glycoprotein binding. *J Biol Chem* **1997**, 272, (2), 1248-55.

128. Yang, L. J.; Lorenzini, I.; Vajn, K.; Mountney, A.; Schramm, L. P.; Schnaar, R. L., Sialidase enhances spinal axon outgrowth in vivo. *Proc Natl Acad Sci U S A* **2006**, 103, (29), 11057-62.
129. Vyas, A. A.; Blixt, O.; Paulson, J. C.; Schnaar, R. L., Potent glycan inhibitors of myelin-associated glycoprotein enhance axon outgrowth in vitro. *J Biol Chem* **2005**, 280, (16), 16305-10.
130. Ito, H.; Ishida, H.; Collins, B. E.; Fromholt, S. E.; Schnaar, R. L.; Kiso, M., Systematic synthesis and MAG-binding activity of novel sulfated GM1b analogues as mimics of Chol-1 (alpha-series) gangliosides: highly active ligands for neural siglecs. *Carbohydr Res* **2003**, 338, (16), 1621-39.
131. Schwardt, O.; Gäthje, H.; Vedani, A.; Mesch, S.; Gao, G. P.; Spreafico, M.; von Orelli, J.; Kelm, S.; Ernst, B., Examination of the biological role of the alpha(2-->6)-linked sialic acid in gangliosides binding to the myelin-associated glycoprotein (MAG). *J Med Chem* **2009**, 52, (4), 989-1004.
132. Bhunia, A.; Schwardt, O.; Gäthje, H.; Gao, G. P.; Kelm, S.; Benie, A. J.; Hricovini, M.; Peters, T.; Ernst, B., Consistent bioactive conformation of the Neu5Acalpha(2-->3)Gal epitope upon lectin binding. *Chembiochem* **2008**, 9, (18), 2941-5.
133. Shin, S. Y.; Gäthje, H.; Schwardt, O.; Gao, G. P.; Ernst, B.; Kelm, S.; Meyer, B., Binding epitopes of gangliosides to their neuronal receptor, myelin-associated glycoprotein, from saturation transfer difference NMR. *Chembiochem* **2008**, 9, (18), 2946-9.
134. Schwizer, D.; Gäthje, H.; Kelm, S.; Porro, M.; Schwardt, O.; Ernst, B., Antagonists of the myelin-associated glycoprotein: a new class of tetrasaccharide mimics. *Bioorg Med Chem* **2006**, 14, (14), 4944-57.
135. Blixt, O.; Collins, B. E.; van den Nieuwenhof, I. M.; Crocker, P. R.; Paulson, J. C., Sialoside specificity of the siglec family assessed using novel multivalent probes: identification of potent inhibitors of myelin-associated glycoprotein. *J Biol Chem* **2003**, 278, (33), 31007-19.
136. Shelke, S. V.; Gao, G. P.; Mesch, S.; Gäthje, H.; Kelm, S.; Schwardt, O.; Ernst, B., Synthesis of sialic acid derivatives as ligands for the myelin-associated glycoprotein (MAG). *Bioorg Med Chem* **2007**, 15, (14), 4951-65.
137. Kelm, S.; Brossmer, R.; Isecke, R.; Gross, H. J.; Strenge, K.; Schauer, R., Functional groups of sialic acids involved in binding to siglecs (sialoadhesins) deduced from interactions with synthetic analogues. *Eur J Biochem* **1998**, 255, (3), 663-72.
138. Strenge, K.; Schauer, R.; Bovin, N.; Hasegawa, A.; Ishida, H.; Kiso, M.; Kelm, S., Glycan specificity of myelin-associated glycoprotein and sialoadhesin deduced from interactions with synthetic oligosaccharides. *Eur J Biochem* **1998**, 258, (2), 677-85.
139. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J., J.A.; Vreven, T.; Kudin, T. K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.;

- Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Adamo, C.; Jaramillo, C.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A., Gaussian 2003. In J.A. Gaussian, Inc: Pittsburgh PA, 2003.
140. Iler, C.; Plesset, M. S., Note on an Approximation Treatment for Many-Electron Systems. *Physical Review* **1934**, 46, (7), 618.
 141. Head-Gordon, M.; Pople, J. A.; Frisch, M. J., MP2 energy evaluation by direct methods. *Chem. Phys. Lett.* **1988**, 153, (6), 503-506.
 142. Simon, S.; Duran, M.; Dannenberg, J. J., How does basis set superposition error change the potential surfaces for hydrogen-bonded dimers? *J. Chem. Phys* **1996**, 105, 11024.
 143. Boys, S. F.; Bernardi, F., The calculation of small molecular interactions by the differences of separate total energies. Some procedures with reduced errors *Mol. Phys.* **1970**, (19), 553.
 144. Vedani, A.; Huhta, D. W., Algorithm for the systematic solvation of proteins based on the directionality of hydrogen bonds. *Journal of the American Chemical Society* **1991**, 113, (15), 5860-5862.
 145. Lu, Y.; Wang, Y.; Zhu, W., Nonbonding interactions of organic halogens in biological systems: implications for drug discovery and biomolecular design. *Phys Chem Chem Phys* **2010**, 12, (18), 4543-51.
 146. Taylor, R.; Kennard, O., Crystallographic Evidence for the Existence of C-H...O, C-H...N, and C-H...Cl Hydrogen Bonds *J. Am. Chem. Soc* **1982**, 104, 5063-5070.
 147. Auffinger, P.; Hays, F. A.; Westhof, E.; Ho, P. S., Halogen bonds in biological molecules. *Proc Natl Acad Sci U S A* **2004**, 101, (48), 16789-94.
 148. Vedani, A.; Huhta, D. W., A NEW FORCE-FIELD FOR MODELING METALLOPROTEINS. *Journal of the American Chemical Society* **1990**, 112, (12), 4759-4767.
 149. Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S.; Weiner, P., A NEW FORCE-FIELD FOR MOLECULAR MECHANICAL SIMULATION OF NUCLEIC-ACIDS AND PROTEINS. *Journal of the American Chemical Society* **1984**, 106, (3), 765-784.
 150. <http://www.schrodinger.com>
 151. Smiesko, M., In preparation.

152. Rossato, G.; Ernst, B.; Smiesko, M.; Spreafico, M.; Vedani, A., Probing Small-Molecule Binding to Cytochrome P450 2D6 and 2C9: An In Silico Protocol for Generating Toxicity Alerts. *ChemMedChem* **2010**, 5, (12), 2088-101.
153. Rowland, P.; Blaney, F. E.; Smyth, M. G.; Jones, J. J.; Leydon, V. R.; Oxbrow, A. K.; Lewis, C. J.; Tennant, M. G.; Modi, S.; Eggleston, D. S.; Chenery, R. J.; Bridges, A. M., Crystal structure of human cytochrome P450 2D6. *Journal of Biological Chemistry* **2006**, 281, (11), 7614-7622.
154. Wester, M. R.; Yano, J. K.; Schoch, G. A.; Yang, C.; Griffin, K. J.; Stout, C. D.; Johnson, E. F., The structure of human cytochrome P4502C9 complexed with flurbiprofen at 2.0-angstrom resolution. *Journal of Biological Chemistry* **2004**, 279, (34), 35630-35637.
155. Zbinden, P.; Dobler, M.; Folkers, G.; Vedani, A., PrGen: Pseudoreceptor modeling using receptor-mediated ligand alignment and pharmacophore equilibration. *Quantitative Structure-Activity Relationships* **1998**, 17, (2), 122-130.
156. Bowers, K. J.; Chow, E.; Huageng, X.; Dror, R. O.; Eastwood, M. P.; Gregersen, B. A.; Klepeis, J. L.; Kolossvary, I.; Moraes, M. A.; Sacerdoti, F. D.; Salmon, J. K.; Shan, Y.; Shaw, D. E., Scalable algorithms for molecular dynamics simulations on commodity clusters. *SC 2006 Proceedings Supercomputing 2006* **2006**, 13 pp.ICD-ROM.
157. McLachlan, A. D., RAPID COMPARISON OF PROTEIN STRUCTURES. *Acta Crystallographica Section A* **1982**, 38, (NOV), 871-873.
158. Jorgensen, W. L., The OPLS Potential Functions for Proteins. Energy Minimization for Crystals of Cyclic Peptides and Crambin. In Tirado-Rives, J., Ed. *J. Am. Chem. Soc.*, 1988; Vol. 110, pp 1657-1666.
159. Cramer, C. J.; Truhlar, D. G., AM1-SM2 AND PM3-SM3 PARAMETERIZED SCF SOLVATION MODELS FOR FREE-ENERGIES IN AQUEOUS-SOLUTION. *Journal of Computer-Aided Molecular Design* **1992**, 6, (6), 629-666.
160. Lill, M. A.; Dobler, M.; Vedani, A., Prediction of small-molecule binding to cytochrome P450 3A4: Flexible docking combined with multidimensional QSAR. *Chemmedchem* **2006**, 1, (1), 73-81.
161. Kumar, V.; Rock, D. A.; Warren, C. J.; Tracy, T. S.; Wahlstrom, J. L., Enzyme source effects on CYP2C9 kinetics and inhibition. *Drug Metabolism and Disposition* **2006**, 34, (11), 1903-1908.
162. Paine, M.; McLaughlin, L.; Flanagan, J.; Kemp, C.; Sutcliffe, M.; Roberts, G.; Wolf, C., Residues glutamate 216 and aspartate 301 are key determinants of substrate specificity and product regioselectivity in cytochrome P450 2D6. *J Biol Chem* **2003**, 278, (6), 4021-7.
163. Flanagan, J.; Maréchal, J.; Ward, R.; Kemp, C.; McLaughlin, L.; Sutcliffe, M.; Roberts, G.; Paine, M.; Wolf, C., Phe120 contributes to the regioselectivity of cytochrome P450 2D6: mutation leads to the formation of a novel dextromethorphan metabolite. *Biochem J* **2004**, 380, (Pt 2), 353-60.

164. Takanashi, K.; Tainaka, H.; Kobayashi, K.; Yasumori, T.; Hosakawa, M.; Chiba, K., CYP2C9 Ile359 and Leu359 variants: enzyme kinetic study with seven substrates. *Pharmacogenetics* **2000**, 10, (2), 95-104.
165. Flanagan, J.; McLaughlin, L.; Paine, M.; Sutcliffe, M.; Roberts, G.; Wolf, C., Role of conserved Asp293 of cytochrome P450 2C9 in substrate recognition and catalytic activity. *Biochem J* **2003**, 370, (Pt 3), 921-6.
166. Melet, A.; Assrir, N.; Jean, P.; Pilar Lopez-Garcia, M.; Marques-Soares, C.; Jaouen, M.; Dansette, P.; Sari, M.; Mansuy, D., Substrate selectivity of human cytochrome P450 2C9: importance of residues 476, 365, and 114 in recognition of diclofenac and sulfaphenazole and in mechanism-based inactivation by tienilic acid. *Arch Biochem Biophys* **2003**, 409, (1), 80-91.
167. Vedani, A.; Dobler, M.; Lill, M., Combining protein modeling and 6D-QSAR. Simulating the binding of structurally diverse ligands to the estrogen receptor. *J Med Chem* **2005**, 48, (11), 3700-3.
168. Vedani, A.; Smiesko, M., *Structure-based computational pharmacology and toxicology*. In *Computational Structural Biology*, Press, I. C., Ed. Eds: Schwede, S. and Peitsch, M.: London, **2008**; pp 549-572.
169. May, A. P.; Robinson, R. C.; Vinson, M.; Crocker, P. R.; Jones, E. Y., Crystal structure of the N-terminal domain of sialoadhesin in complex with 3' sialyllactose at 1.85 Å resolution. *Mol Cell* **1998**, 1, (5), 719-28.
170. Shi, J.; Blundell, T. L.; Mizuguchi, K., FUGUE: sequence-structure homology recognition using environment-specific substitution tables and structure-dependent gap penalties. *J Mol Biol* **2001**, 310, (1), 243-57.
171. Deane, C. M.; Kaas, Q.; Blundell, T. L., SCORE: predicting the core of protein models. *Bioinformatics* **2001**, 17, (6), 541-50.
172. Deane, C. M.; Blundell, T. L., CODA: a combined algorithm for predicting the structurally variable regions of protein models. *Protein Sci* **2001**, 10, (3), 599-612.
173. TRIPOS. <http://tripos.com/>
174. Thompson, J. D.; Higgins, D. G.; Gibson, T. J., CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **1994**, 22, (22), 4673-80.
175. Case, D. A.; Pearlman, D. A.; Caldwell, J. W. In *AMBER 7*, University of California: San Francisco, 2002; University of California: San Francisco, 2002.
176. Jarzynski, C., Equilibrium free-energy differences from nonequilibrium measurements: A master-equation approach. *Physical Review E* **1997**, 56, (5), 5018-5035.
177. Lamb, M. L.; Jorgensen, W. L., Computational approaches to molecular recognition. *Curr Opin Chem Biol* **1997**, 1, (4), 449-57.
178. Kollman, P., Free energy calculations: Applications to chemical and biochemical phenomena. *Chem. Rev.* **1993**, 93, (7), 2395-2417.

179. Zwanzig, R. W., High-Temperature Equation of State by a Perturbation Method. I. Nonpolar Gases. *Journal of Chemical Physics* **1954**, 22, 1420-1426.
180. Klemperer, W., The Potential to Surprise. *Nature* **1993**, 362, 698.
181. DePristo, M.; de Bakker, P.; Blundell, T., Heterogeneity and inaccuracy in protein structures solved by X-ray crystallography. *Structure* **2004**, 12, (5), 831-8.
182. Henrik, C.; Görbitz, C. H.; Etter, M. C., Hydrogen bonds to carboxylate groups. The question of three-centre interactions. *Journal of the Chemical Society, Perkin Transactions 2* **1992**, 131-135.
183. Bohm, H. J.; Schneider, G., *Protein-Ligand Interactions from Molecular Recognition to Drug Design*. Wiley-VCH Verlag GmbH and Co.: Weinheim, Germany., 2003.
184. Baron, R.; Setny, P.; McCammon, J. A., Water in cavity-ligand recognition. *J Am Chem Soc* **2010**, 132, (34), 12091-7.
185. Plumridge, T. H.; Waigh, R. D., Water structure theory and some implications for drug design. *J Pharm Pharmacol* **2002**, 54, (9), 1155-79.
186. NIST Chemistry WebBook. webbook.nist.gov
187. Laurence, C.; Brameld, K.; Graton, J.; Le Questel, J.; Renault, E., The pK(BHX) database: toward a better understanding of hydrogen-bond basicity for medicinal chemists. *J Med Chem* **2009**, 52, (14), 4073-86.
188. Gelin, B. R.; Karplus, M., Side-chain torsional potentials: effect of dipeptide, protein, and solvent environment. *Biochemistry* **1979**, 18, (7), 1256-68.
189. Seifert, A.; Tatzel, S.; Schmid, R. D.; Pleiss, J., Multiple molecular dynamics simulations of human P450 monooxygenase CYP2C9: The molecular basis of substrate binding and regioselectivity toward warfarin. *Proteins-Structure Function and Bioinformatics* **2006**, 64, (1), 147-155.
190. de Groot, M.; Ackland, M.; Horne, V.; Alex, A.; Jones, B., Novel approach to predicting P450-mediated drug metabolism: development of a combined protein and pharmacophore model for CYP2D6. *J Med Chem* **1999**, 42, (9), 1515-24.
191. de Groot, M. J.; Ackland, M. J.; Horne, V. A.; Alex, A. A.; Jones, B. C., A novel approach to predicting P450 mediated drug metabolism. CYP2D6 catalyzed N-dealkylation reactions and qualitative metabolite predictions using a combined protein and pharmacophore model for CYP2D6. *Journal of Medicinal Chemistry* **1999**, 42, (20), 4062-4070.
192. Zhang, D.; Freeman, J.; Sutherland, J.; Walker, A.; Yang, Y.; Cerniglia, C., Biotransformation of chlorpromazine and methdilazine by *Cunninghamella elegans*. *Appl Environ Microbiol* **1996**, 62, (3), 798-803.
193. Hanna, I.; Kim, M.; Guengerich, F., Heterologous expression of cytochrome P450 2D6 mutants, electron transfer, and catalysis of bufuralol hydroxylation: the role of aspartate 301 in structural integrity. *Arch Biochem Biophys* **2001**, 393, (2), 255-61.

194. M., D.; H., H.; T., G.; D., L.; M., H. R.; N., K. C.; A., K. R.; K., A., Identification of trifluoperazine metabolites using HPLC–NMR of test mixtures. In *Magn. Reson. Chem*, 2000; Vol. 38, pp 951-956.
195. Yamazaki, H.; Shimada, T., Progesterone and testosterone hydroxylation by cytochromes P450 2C19, 2C9, and 3A4 in human liver microsomes. *Arch Biochem Biophys* **1997**, 346, (1), 161-9.
196. Shen, H.; He, M.; Liu, H.; Wrighton, S.; Wang, L.; Guo, B.; Li, C., Comparative metabolic capabilities and inhibitory profiles of CYP2D6.1, CYP2D6.10, and CYP2D6.17. *Drug Metab Dispos* **2007**, 35, (8), 1292-300.
197. Mannens, G.; Snel, C.; Hendrickx, J.; Verhaeghe, T.; Le Jeune, L.; Bode, W.; van Beijsterveldt, L.; Lavrijsen, K.; Leempoels, J.; Van Osselaer, N.; Van Peer, A.; Meuldermans, W., The metabolism and excretion of galantamine in rats, dogs, and humans. *Drug Metab Dispos* **2002**, 30, (5), 553-63.
198. Nielsen, K.; Brøsen, K.; Gram, L., Steady-state plasma levels of clomipramine and its metabolites: impact of the sparteine/debrisoquine oxidation polymorphism. Danish University Antidepressant Group. *Eur J Clin Pharmacol* **1992**, 43, (4), 405-11.
199. Zamora, I.; Afzelius, L.; Cruciani, G., Predicting drug metabolism: a site of metabolism prediction tool applied to the cytochrome P450 2C9. *J Med Chem* **2003**, 46, (12), 2313-24.
200. Wold, S., Methods and Principles in Medicinal Chemistry. In Eriksson, L. v. d. W., H., Ed. VCH: Chemometric Methods in Molecular Design, 1995; pp 309–318.
201. Hamelin, B.; Bouayad, A.; Drolet, B.; Gravel, A.; Turgeon, J., In vitro characterization of cytochrome P450 2D6 inhibition by classic histamine H1 receptor antagonists. *Drug Metab Dispos* **1998**, 26, (6), 536-9.
202. Strobl, G.; von Kruedener, S.; Stöckigt, J.; Guengerich, F.; Wolff, T., Development of a pharmacophore for inhibition of human liver cytochrome P-450 2D6: molecular modeling and inhibition studies. *J Med Chem* **1993**, 36, (9), 1136-45.
203. Rao, S.; Aoyama, R.; Schrag, M.; Trager, W.; Rettie, A.; Jones, J., A refined 3-dimensional QSAR of cytochrome P450 2C9: computational predictions of drug interactions. *J Med Chem* **2000**, 43, (15), 2789-96.
204. Lill, M. A.; Winiger, F.; Vedani, A.; Ernst, B., Impact of induced fit on ligand binding to the androgen receptor: A multidimensional QSAR study to predict endocrine-disrupting effects of environmental chemicals. *Journal of Medicinal Chemistry* **2005**, 48, (18), 5666-5674.
205. Vedani, A.; Dobler, M.; Lill, M., The challenge of predicting drug toxicity in silico. *Basic Clin Pharmacol Toxicol* **2006**, 99, (3), 195-208.
206. Spreafico, M.; Ernst, B.; Lill, M.; Smiesko, M.; Vedani, A., Mixed-model QSAR at the glucocorticoid receptor: predicting the binding mode and affinity of psychotropic drugs. *ChemMedChem* **2009**, 4, (1), 100-9.

207. Spreafico, M.; Smiesko, M.; Peristera, O.; Rossato, G.; Vedani, A., Probing small-molecule binding to the Liver-X receptor: A mixed-model QSAR study. In *Mol. Inf.*, 2010; Vol. 1, pp 27-36.
208. Peristera, O.; Spreafico, M.; Smiesko, M.; Ernst, B.; Vedani, A., Mixed-model QSAR at the human mineralocorticoid receptor: predicting binding mode and affinity of anabolic steroids. *Toxicol Lett* **2009**, 189, (3), 219-24.
209. Vedani, A.; Descloux, A.; Spreafico, M.; Ernst, B., Predicting the toxic potential of drugs and chemicals in silico: a model for the peroxisome proliferator-activated receptor gamma (PPAR gamma). *Toxicol Lett* **2007**, 173, (1), 17-23.
210. Vedani, A.; Zumstein, M.; Lill, M.; Ernst, B., Simulating alpha/beta selectivity at the human thyroid hormone receptor: consensus scoring using multidimensional QSAR. *ChemMedChem* **2007**, 2, (1), 78-87.
211. Lin, L.; Di Stefano, E.; Schmitz, D.; Hsu, L.; Ellis, S.; Lennard, M.; Tucker, G.; Cho, A., Oxidation of methamphetamine and methylenedioxymethamphetamine by CYP2D6. *Drug Metab Dispos* **1997**, 25, (9), 1059-64.
212. Preskorn, S., Reproducibility of the in vivo effect of the selective serotonin reuptake inhibitors on the in vivo function of cytochrome P450 2D6: an update (part I). *J Psychiatr Pract* **2003**, 9, (2), 150-8.
213. Li, X.; Andersson, T.; Ahlström, M.; Weidolf, L., Comparison of inhibitory effects of the proton pump-inhibiting drugs omeprazole, esomeprazole, lansoprazole, pantoprazole, and rabeprazole on human cytochrome P450 activities. *Drug Metab Dispos* **2004**, 32, (8), 821-7.
214. Sinnokrot, M. O.; Sherrill, C. D., Substituent effects in pi-pi interactions: sandwich and T-shaped configurations. *J Am Chem Soc* **2004**, 126, (24), 7690-7.
215. Kelm, S.; Brossmer, R. Neuraminic acid derivatives for use as Siglec inhibitors. 2003.

Appendix 1: Supporting Information of “Water in molecular simulations and drug design”

Supporting Information

Title:

AcquaAlta: a directional approach to the solvation of ligand–protein complexes

Authors:

Gianluca Rossato, Beat Ernst, Angelo Vedani, and Martin Smiesko

Institute of Molecular Pharmacy, Pharmazenter,
University of Basel,
Klingelbergstr. 50,
4056 Basel, Switzerland

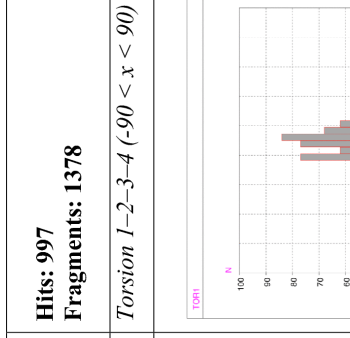
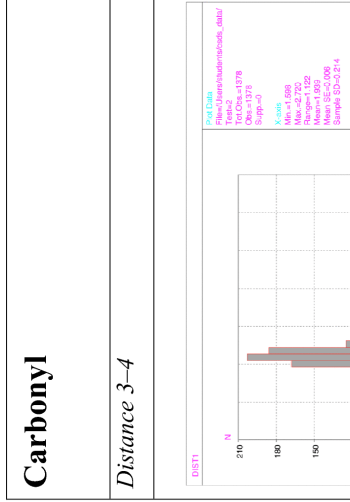
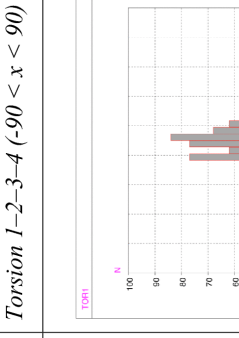
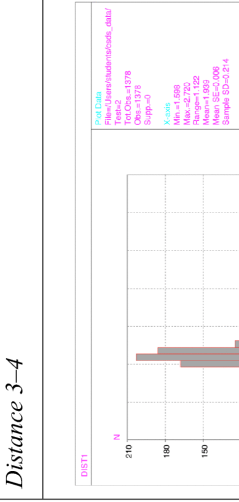
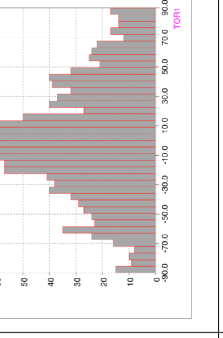
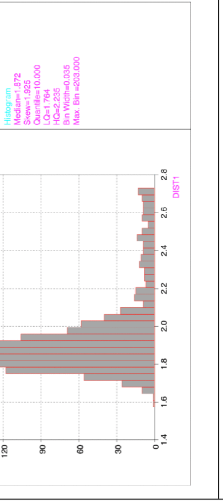
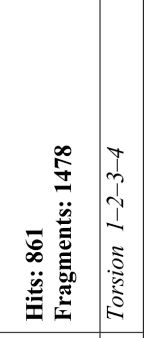
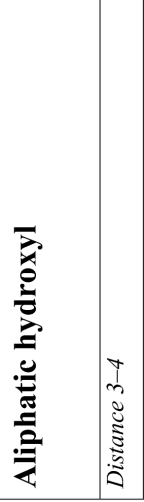
List of Figures and Tables:

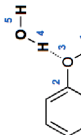
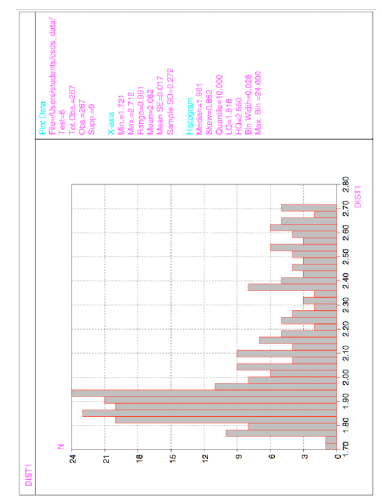
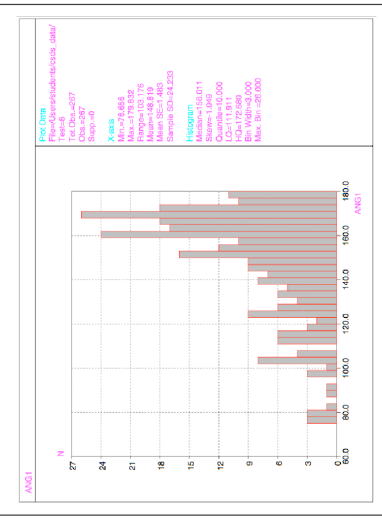
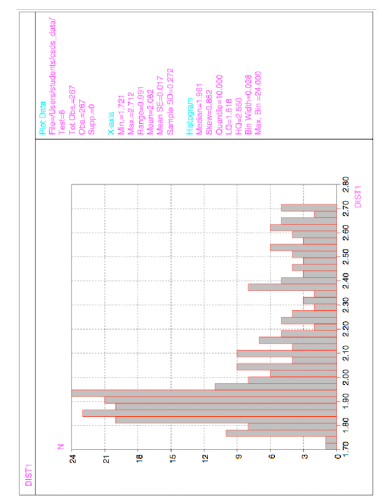
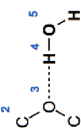
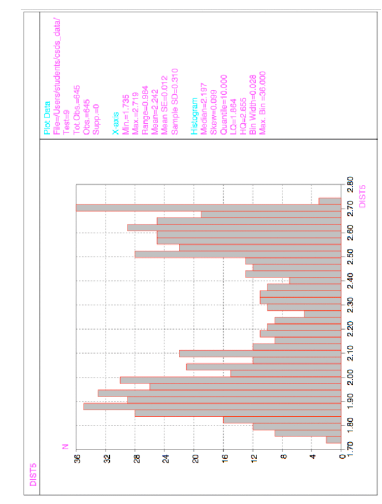

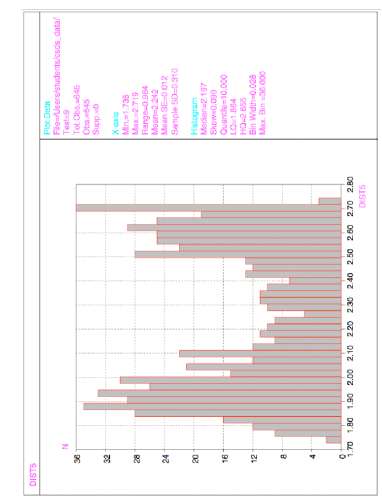
Figure S1: Statistics for distances, angles and torsions from the CSD search.

Table S1a: Geometries and interaction energies as from *ab initio* calculations for hydrogen-bond acceptors. Listed also atom hybridization, formal charge, and number of lone pairs. The propensity to hydration (HP) is ranked going from easier to more difficult probability of hydration.

Table S1b: Geometries and interaction energies as from *ab initio* calculations for hydrogen-bond donors. Listed also atom hybridization, formal charge, and number of lone pairs. The propensity to hydration (HP) is ranked going from easier to more difficult probability of hydration.

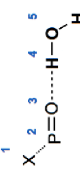
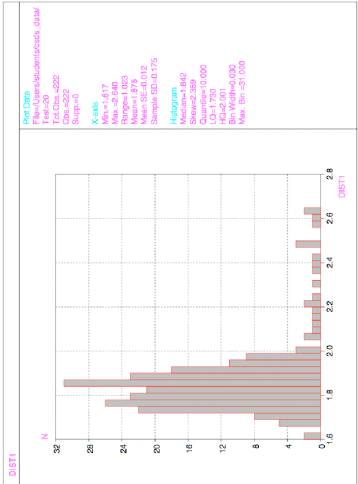
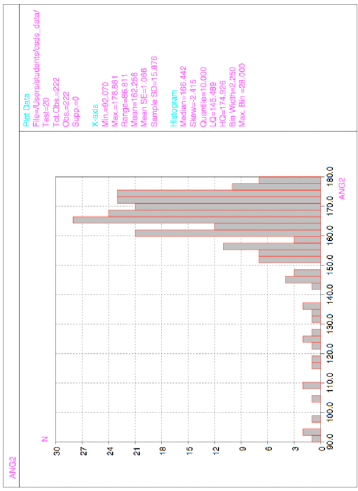
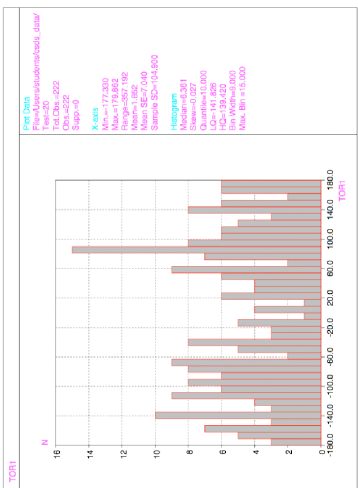
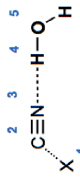
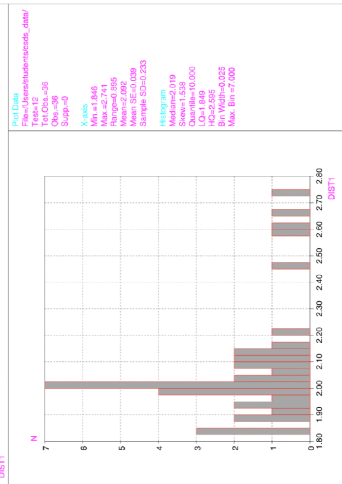
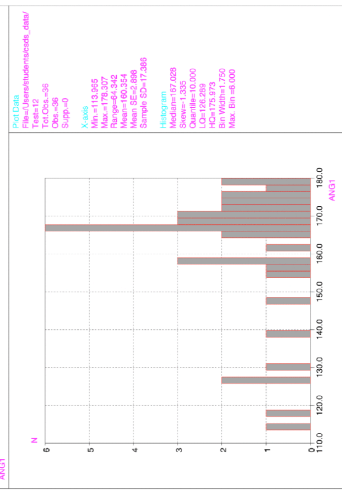
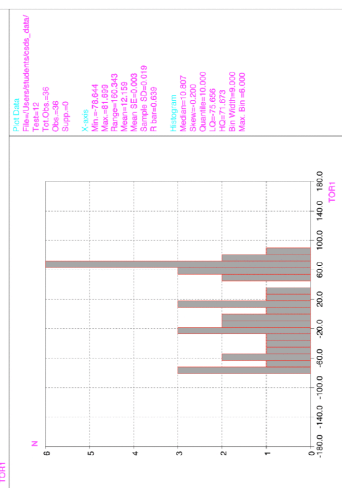
Table S2: Comparison of values from queries using hydroxyl considered as a fragment and explicit water molecules.


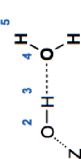

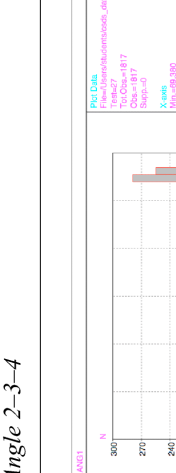
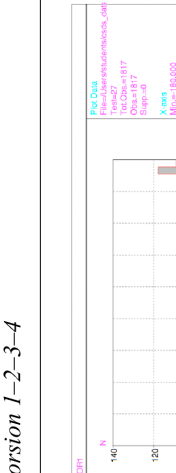

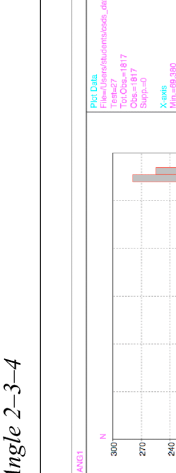
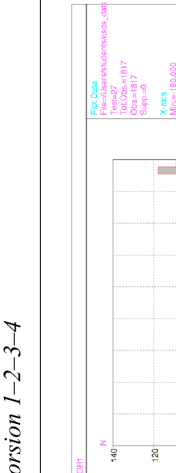
<p>Carbonyl</p>  <p>X = Any kind of atom</p>	<p>Hits: 997 Fragments: 1378</p> <p><i>Torsion 1-2-3-4 (-90 < x < 90)</i></p>	<p>Distance 3-4</p>  <p> <small> Plot Data File: Users\shimadzu\..._data1 Tol.Cnts=1378 Obs=1478 Steps=2 X-axis Min=1.688 Max=2.750 Mean=1.922 Median=1.899 Sample S.D.=0.214 Histogram Median=1.872 Skewness=0.000 Kurtosis=0.000 LD=1.794 HD=1.944 Bin Width=0.035 Bin Volumes=0.000 Max. Bin =025.000 </small> </p>
<p>Aliphatic hydroxyl</p> 	<p>Hits: 861 Fragments: 1478</p> <p><i>Torsion 1-2-3-4</i></p>	<p>Distance 3-4-5</p>  <p> <small> Plot Data File: Users\shimadzu\..._data1 Tol.Cnts=1378 Obs=1478 Steps=2 X-axis Min=77.628 Max=173.904 Mean=100.018 Median=100.013 Sample S.D.=17.467 Histogram Median=106.724 Skewness=0.000 Kurtosis=0.000 LD=104.105 HD=109.930 Bin Width=2.000 Bin Volumes=0.000 Max. Bin =18.000 </small> </p>
<p>Carbonyl</p> 	<p>Hits: 997 Fragments: 1378</p> <p><i>Torsion 1-2-3-4 (-90 < x < 90)</i></p>	<p>Distance 3-4</p>  <p> <small> Plot Data File: Users\shimadzu\..._data1 Tol.Cnts=1378 Obs=1478 Steps=2 X-axis Min=1.688 Max=2.750 Mean=1.922 Median=1.899 Sample S.D.=0.214 Histogram Median=1.872 Skewness=0.000 Kurtosis=0.000 LD=1.794 HD=1.944 Bin Width=0.035 Bin Volumes=0.000 Max. Bin =025.000 </small> </p>
<p>Aliphatic hydroxyl</p> 	<p>Hits: 861 Fragments: 1478</p> <p><i>Torsion 1-2-3-4</i></p>	<p>Distance 3-4-5</p>  <p> <small> Plot Data File: Users\shimadzu\..._data1 Tol.Cnts=1378 Obs=1478 Steps=2 X-axis Min=77.628 Max=173.904 Mean=100.018 Median=100.013 Sample S.D.=17.467 Histogram Median=106.724 Skewness=0.000 Kurtosis=0.000 LD=104.105 HD=109.930 Bin Width=2.000 Bin Volumes=0.000 Max. Bin =18.000 </small> </p>

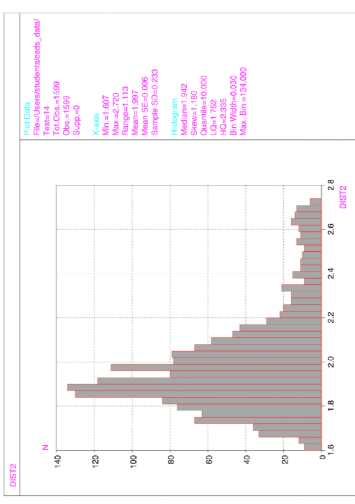
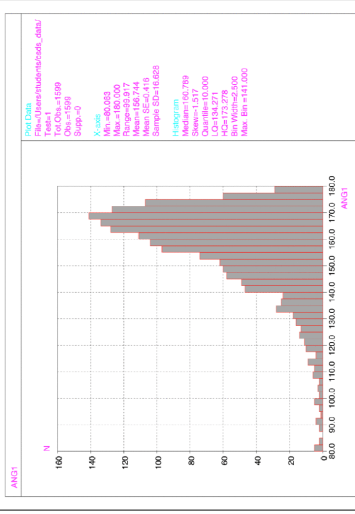
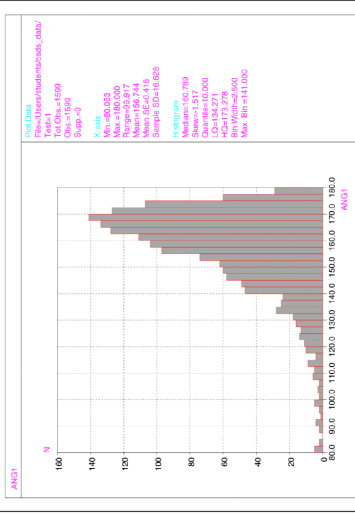
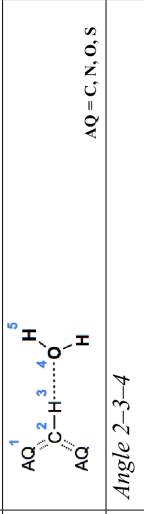
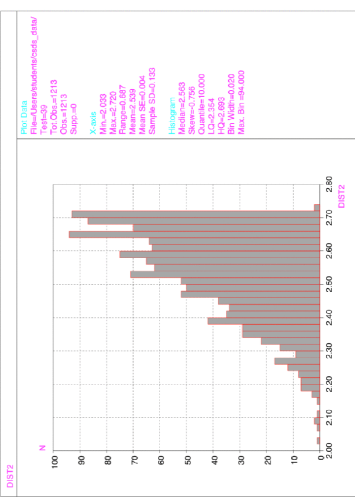
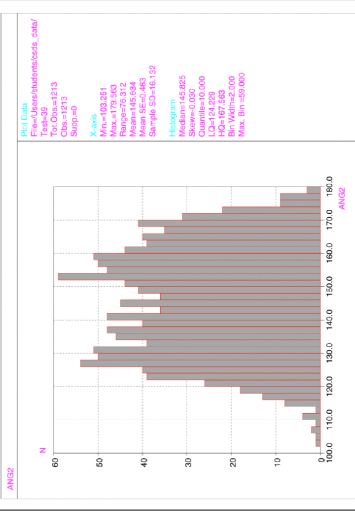
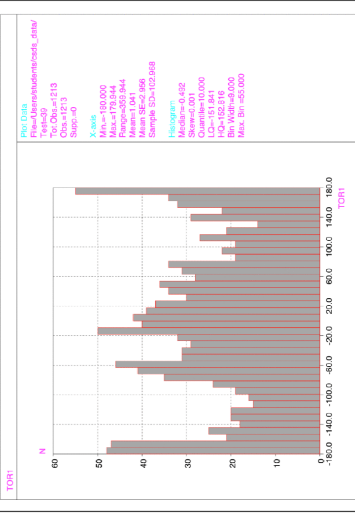
Phenolic hydroxyl		<p>Hits: 168 Fragments: 267</p> <p><i>Torsion 1-2-3-4</i></p>
<p>Distance 3-4</p> 	<p>Angle 3-4-5</p> 	<p>Distance 3-4</p> 
<p>Ether</p> <p><i>Distance 3-4</i></p>	 <p>Angle 3-4-5</p>	<p>Hits: 375 Fragments: 645</p> <p><i>Torsion 1-2-3-4</i></p>
<p>Distance 3-4</p> 	<p>Angle 3-4-5</p> 	<p>Distance 3-4</p> 

<p>Imine</p> <p><i>Distance 3-4</i></p>	<p>Z_2 Z_1 Z_3 Z_4 Z_5</p> <p>Z= No hydrogen- (***) neutral</p> <p><i>Angle 3-4-5</i></p>	<p>Hits: 451</p> <p>Fragments: 671</p> <p><i>Torsion 1-2-3-4</i></p>
<p>Dist Data File: L:\Users\stamir\stamir\data\TDCData\#71 Date: 1/1/2014 Step: #0</p> <p>Stats Min: 1.728 Max: 2.811 Mean: 2.112 Median: 2.098 StdDev: 0.208</p> <p>Histogram Median: 1.972 StdDev: 0.208 Char: 100.000 LD: 1.000 Bin Width: 0.100 Bin: #0: 0.000 Max. Bin: #10: 0.000</p>	<p>Ang Data File: L:\Users\stamir\stamir\data\TDCData\#71 Date: 1/1/2014 Step: #0</p> <p>Stats Min: 83.733 Max: 177.209 Mean: 117.388 Median: 117.560 StdDev: 17.560</p> <p>Histogram Median: 113.894 StdDev: 17.560 Char: 100.000 LD: 1.000 Bin Width: 2.500 Bin: #0: 0.000 Max. Bin: #68: 0.000</p>	<p>Phi Data File: L:\Users\stamir\stamir\data\TDCData\#71 Date: 1/1/2014 Step: #0</p> <p>Stats Min: 179.822 Max: 179.822 Mean: 179.822 Median: 179.822 StdDev: 0.000</p> <p>Histogram Median: 179.800 StdDev: 0.000 Char: 100.000 LD: 1.000 Bin Width: 0.000 Bin: #0: 0.000 Max. Bin: #17: 0.000</p>
<p>Thioether (R-S-R)</p> <p><i>Distance 3-4</i></p>	<p>Z_2 Z_3 Z_4 Z_5</p> <p>Z= No hydrogen - (***) max two bonds</p> <p><i>Angle 3-4-5</i></p>	<p>Hits: 11</p> <p>Fragments: 14</p> <p><i>Torsion 1-2-3-4</i></p>
<p>Dist Data File: L:\Users\stamir\stamir\data\TDCData\#14 Date: 1/1/2014 Step: #0</p> <p>Stats Min: 2.416 Max: 2.934 Mean: 2.500 Median: 2.500 StdDev: 0.250</p> <p>Histogram Median: 2.500 StdDev: 0.250 Char: 100.000 LD: 1.000 Bin Width: 0.015 Bin: #0: 0.000 Max. Bin: #3: 0.000</p>	<p>Ang Data File: L:\Users\stamir\stamir\data\TDCData\#14 Date: 1/1/2014 Step: #0</p> <p>Stats Min: 87.777 Max: 173.851 Mean: 130.819 Median: 127.000 StdDev: 32.819</p> <p>Histogram Median: 89.851 StdDev: 32.819 Char: 100.000 LD: 1.000 Bin Width: 2.000 Bin: #0: 0.000 Max. Bin: #2: 0.000</p>	<p>Phi Data File: L:\Users\stamir\stamir\data\TDCData\#14 Date: 1/1/2014 Step: #0</p> <p>Stats Min: 179.307 Max: 179.431 Mean: 179.369 Median: 179.308 StdDev: 0.067</p> <p>Histogram Median: 179.309 StdDev: 0.067 Char: 100.000 LD: 1.000 Bin Width: 0.010 Bin: #0: 0.000 Max. Bin: #2: 0.000</p>

<p>Thioketon</p> <p><i>Distance 3-4</i></p>	<p>Hits: 59 Fragments: 76</p> <p><i>Torsion 1-2-3-4</i></p>	<p>Hits: 139 Fragments: 271</p> <p><i>Torsion 1-2-3-4</i></p>
<p>Sulfoxide / Sulfone</p> <p><i>Distance 3-4</i></p>	<p>Hits: 59 Fragments: 76</p> <p><i>Torsion 1-2-3-4</i></p>	<p>Hits: 139 Fragments: 271</p> <p><i>Torsion 1-2-3-4</i></p>
<p>Thioketon</p> <p><i>Distance 3-4</i></p>	<p>Hits: 59 Fragments: 76</p> <p><i>Torsion 1-2-3-4</i></p>	<p>Hits: 139 Fragments: 271</p> <p><i>Torsion 1-2-3-4</i></p>
<p>Sulfoxide / Sulfone</p> <p><i>Distance 3-4</i></p>	<p>Hits: 59 Fragments: 76</p> <p><i>Torsion 1-2-3-4</i></p>	<p>Hits: 139 Fragments: 271</p> <p><i>Torsion 1-2-3-4</i></p>

<p>Phosphine oxide / Phosphone</p> <p><i>Distance 3-4</i></p>	<p>X= Any kind of atom</p> <p><i>Angle 3-4-5</i></p> 	<p>Hits: 133 Fragments: 271</p> <p><i>Torsion 1-2-3-4</i></p>
 <p>Dist Data File: J:\chem\chem\mcs\mcs_data\ Title: Dist_3-4 TOL Obs: 222 Obs: 222 Suppl: 0</p> <p>X Stats Min: 1.817 Max: 2.711 Mean: 1.879 Median: 1.822 Sample SD: 0.179</p> <p>Histogram Median: 1.842 Standard Deviation: 0.180 Quantiles: 0.000 LC: 1.750 Bin Width: 0.200 Bin Width: 0.200 Max. Bin: 0.500</p>	 <p>Ang Data File: J:\chem\chem\mcs\mcs_data\ Title: Ang_3-4-5 TOL Obs: 222 Obs: 222 Suppl: 0</p> <p>X Stats Min: 96.070 Max: 179.111 Mean: 124.256 Median: 121.870 Sample SD: 15.870</p> <p>Histogram Median: 122.442 Standard Deviation: 15.000 Quantiles: 0.000 LC: 115.889 Bin Width: 2.000 Bin Width: 2.000 Max. Bin: 0.500</p>	 <p>Tor Data File: J:\chem\chem\mcs\mcs_data\ Title: Tor_1-2-3-4 TOL Obs: 222 Obs: 222 Suppl: 0</p> <p>X Stats Min: -177.260 Max: 177.260 Mean: 0.000 Median: 0.000 Sample SD: 104.000</p> <p>Histogram Median: 0.000 Standard Deviation: 104.000 Quantiles: 0.000 LC: -18.000 Bin Width: 2.000 Bin Width: 2.000 Max. Bin: 0.500</p>
<p>Nitrile</p> <p><i>Distance 3-4</i></p>	<p>X= Any kind of atom</p> <p><i>Angle 3-4-5</i></p> 	<p>Hits: 30 Fragments: 36</p> <p><i>Torsion 1-2-3-4</i></p>
 <p>Dist Data File: J:\chem\chem\mcs\mcs_data\ Title: Dist_3-4 TOL Obs: 36 Obs: 36 Suppl: 0</p> <p>X Stats Min: 1.845 Max: 2.711 Mean: 2.022 Median: 1.982 Sample SD: 0.229</p> <p>Histogram Median: 2.019 Standard Deviation: 0.229 Quantiles: 0.000 LC: 1.849 Bin Width: 0.050 Bin Width: 0.050 Max. Bin: 0.200</p>	 <p>Ang Data File: J:\chem\chem\mcs\mcs_data\ Title: Ang_3-4-5 TOL Obs: 36 Obs: 36 Suppl: 0</p> <p>X Stats Min: 113.905 Max: 179.307 Mean: 132.243 Median: 130.324 Sample SD: 18.890</p> <p>Histogram Median: 127.029 Standard Deviation: 18.000 Quantiles: 0.000 LC: 116.290 Bin Width: 2.000 Bin Width: 2.000 Max. Bin: 0.500</p>	 <p>Tor Data File: J:\chem\chem\mcs\mcs_data\ Title: Tor_1-2-3-4 TOL Obs: 36 Obs: 36 Suppl: 0</p> <p>X Stats Min: 139.844 Max: -81.820 Mean: 12.243 Median: 12.139 Sample SD: 89.019</p> <p>Histogram Median: 0.000 Standard Deviation: 89.000 Quantiles: 0.000 LC: -75.000 Bin Width: 2.000 Bin Width: 2.000 Max. Bin: 0.500</p>

<p>Aromatic Ring Acceptor</p>		<p>Hits: 40 Fragments: 43</p>
<p><i>Distance 3-4</i></p>	<p><i>Angle 3-4-5</i></p>	<p><i>Torsion 1-2-3-4</i></p>
<p>HEV (OH)</p>	<p><i>Z = Any kind of atom not hydrogen</i></p> 	<p>Hits: 1266 Fragments: 1814</p>
<p><i>Distance 3-4</i></p>	<p><i>Angle 2-3-4</i></p>	<p><i>Torsion 1-2-3-4</i></p>
 <p>Plot Data File: J:/chem/cheminformatics_jmr/ Title: TIC: Obs=43 X-axis: Y-axis: Min=2.342 Max=2.900 Mean=2.509 Median=2.517 StdDev=0.161 Histogram Median=2.574 Skewness=0.000 Kurtosis=0.000 ChiSq=10.000 LD=0.996 Bin Width=0.115 Bin Start=2.300 Max. Bin=240.000</p>	 <p>Plot Data File: J:/chem/cheminformatics_jmr/ Title: TIC: Obs=43 X-axis: Y-axis: Min=98.282 Max=178.000 Mean=113.939 Median=113.851 StdDev=33.877 Histogram Median=98.646 Skewness=0.000 Kurtosis=0.000 ChiSq=10.000 LD=1.179 Bin Width=4.500 Bin Start=0.000 Max. Bin=270.000</p>	 <p>Plot Data File: J:/chem/cheminformatics_jmr/ Title: TIC: Obs=43 X-axis: Y-axis: Min=-109.809 Max=180.000 Mean=7.903 Median=11.151 StdDev=41.817 Histogram Median=76.622 Skewness=0.000 Kurtosis=0.000 ChiSq=10.000 LD=10.096 Bin Width=4.000 Bin Start=-180.000 Max. Bin=10.000</p>
 <p>Plot Data File: J:/chem/cheminformatics_jmr/ Title: TIC: Obs=1817 X-axis: Y-axis: Min=1.487 Max=2.719 Mean=1.736 Median=1.736 StdDev=0.099 Histogram Median=1.791 Skewness=0.000 Kurtosis=0.000 ChiSq=10.000 LD=1.180 Bin Width=0.085 Bin Start=1.400 Max. Bin=240.000</p>	 <p>Plot Data File: J:/chem/cheminformatics_jmr/ Title: TIC: Obs=1817 X-axis: Y-axis: Min=68.380 Max=180.000 Mean=102.702 Median=102.702 StdDev=33.048 Histogram Median=68.697 Skewness=0.000 Kurtosis=0.000 ChiSq=10.000 LD=1.180 Bin Width=3.500 Bin Start=60.000 Max. Bin=270.000</p>	 <p>Plot Data File: J:/chem/cheminformatics_jmr/ Title: TIC: Obs=1817 X-axis: Y-axis: Min=85.000 Max=180.000 Mean=107.000 Median=107.000 StdDev=33.000 Histogram Median=76.000 Skewness=0.000 Kurtosis=0.000 ChiSq=10.000 LD=10.000 Bin Width=3.500 Bin Start=-180.000 Max. Bin=140.000</p>

<p>HEV (NH)</p>	<p>Hits: 1138 Fragments: 1599</p> <p><i>Torsion 1-2-3-4</i></p>	<p>Hits: 758 Fragments: 1213</p> <p><i>Torsion 1-2-3-4</i></p>
<p><i>Distance 3-4</i></p> 	<p><i>Angle 2-3-4</i></p> 	<p><i>Angle 2-3-4</i></p> 
<p>HPV - CH (Caromatic)</p> <p><i>Distance 3-4</i></p>	<p><i>Angle 2-3-4</i></p> 	<p><i>Angle 2-3-4</i></p> <p>AQ = C, N, O, S</p>
		

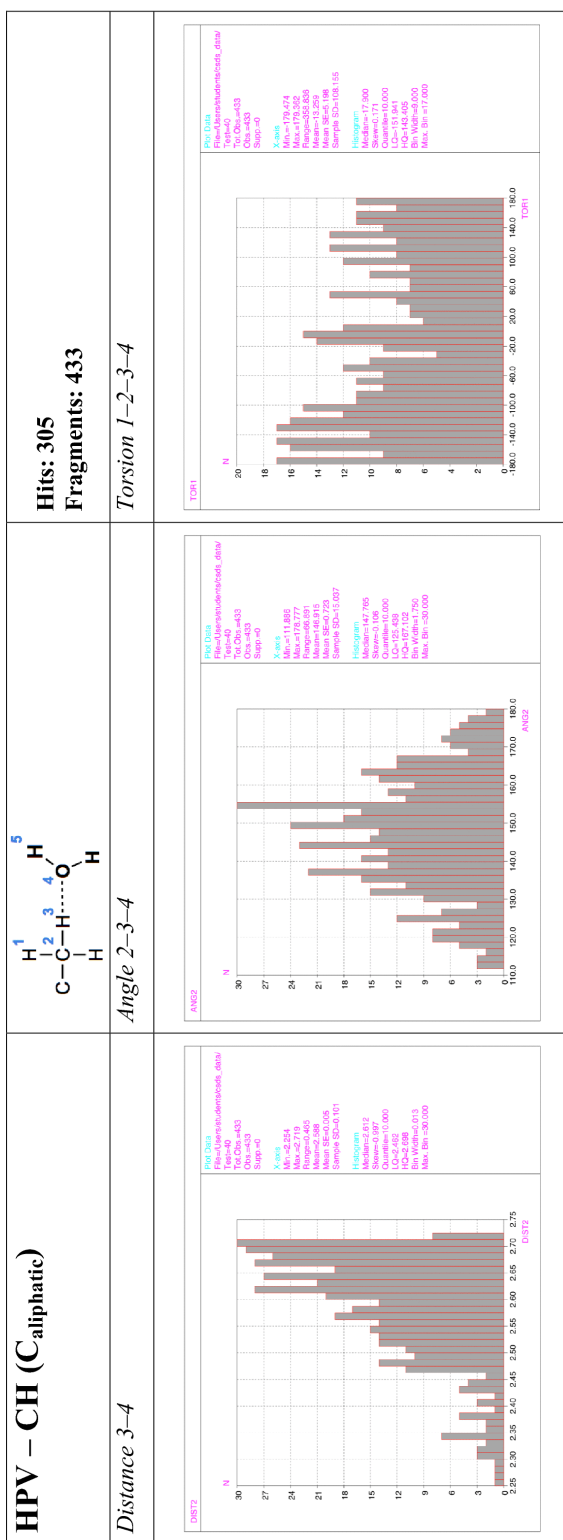


Figure S1: Statistics for distances, angles and torsions from the CSD search.

Table S1a. Geometries and interaction energies as from *ab initio* calculations for hydrogen-bond acceptors. Listed also atom hybridization, formal charge, and number of lone pairs. The propensity to hydration (HP) is ranked going from easier to more difficult probability of hydration.

HP	atom type (typical molecule)	hybridization	ΔE [kcal/mol] distance [Å] angle [°]	formal charge	# of lone pairs
Hydrogen bond acceptors					
1	C=O (acetate, acetone)	sp ²	-14.7 – -5.2 1.73 – 1.94 169.1 – 165.9	-1, 0	2
2	-N= (pyridine)	sp ²	-5.9 1.95 177.3	0	1
3	C=N (acetonitril)	sp	-4.3 2.11 179.9	0	1
4	P=O, S=O, N=O (phosphate, sulfate, nitrobenzene)	sp ²	-7.0 – -3.5 1.85 – 2.12 154.5 – 162.5	0	2
5	-O- (ether, methanol, methylacetate)	sp ³	-5.0 – -3.3 1.86 – 1.97 161.8 – 172.4	0	2
6	-S- (thioether, thiomethanol)	sp ³	-3.8 – -2.9 2.43 – 2.47 145.6 – 151.3	0	2
7	C-X (fluorobenzene)	N.A.	-3.3 – -2.5 2.03 – 2.79 129.8 – 144.9	0	N.A.
8	C _{aromatic} (benzene)	sp ²	-2.2 2.61 170.1	0	N.A.

Table S1b. Geometries and interaction energies as from *ab initio* calculations for hydrogen-bond donors. Listed also atom hybridization, formal charge, and number of lone pairs. The propensity to hydration (HP) is ranked going from easier to more difficult probability of hydration.

HP	atom type (typical molecule)	hybridization	ΔE [kcal/mol] distance [Å] angle [°]	formal charge	# of lone pairs
Hydrogen bond donors					
1	N-H (methylammonium, pyrazol, amide)	N.A.	-17.7 – -4.5 1.70 – 2.03 175.3 – 179.7	1, 0	N.A.
2	O-H (phenol, methanol)	N.A.	-5.9 – 4.3 1.88 – 1.94 177.2 – 180.0	0	N.A.
3	S-H (thiomethanol)	N.A.	-2.22 171.7	0	N.A.
4	C-H (acetylene, benzene, methane)	N.A.	-2.45 – -0.3 2.20 – 2.55 178.6 – 179.5	0	N.A.
5	C-Br (bromobenzene, bromomethane)	N.A.	-0.8 – -0.3 3.13 – 3.19 178.4 – 179.1	0	N.A.
6	C-Cl (chlorobenzene)	N.A.	-0.16 3.11 179.7	0	N.A.

Table S2: Comparison of values from queries using hydroxyl considered as a fragment and explicit water molecules.

Functional groups HBA

Carbonyl OH - FRAGMENT Hits: 8361	Distance (Å)		WATER Hits: 997	Distance (Å)	
	Mean	1.875 ± 0.229		Mean	1.939 ± 0.214
	Median	1.824		Median	1.872
	LQ/HQ	1.661 / 2.115		LQ/HQ	1.764 / 2.235
	Angle (°) Linearity			Angle (°) Linearity	
	Mean	160.138 ± 16.863		Mean	160.019 ± 17.487
	Median	165.363	Median	165.754	
	LQ/HQ	135.988 / 175.608	LQ/HQ	134.125 / 175.145	
Aliphatic OH OH - FRAGMENT Hits: 5469	Distance (Å)		WATER Hits: 861	Distance (Å)	
	Mean	1.957 ± 0.761		Mean	2.031 ± 0.280
	Median	1.849	Median	1.911	
	LQ/HQ	1.725 / 2.448	LQ/HQ	1.780 / 2.537	
	Angle (°) Linearity			Angle (°) Linearity	
	Mean	155.238 ± 20.402		Mean	151.910 ± 23.040
	Median	161.767	Median	159.663	
	LQ/HQ	123.779 / 173.916	LQ/HQ	115.242 / 173.560	
Aromatic hydroxyl OH - FRAGMENT Hits: 946	Distance (Å)		WATER Hits: 168	Distance (Å)	
	Mean	2.022 ± 0.290		Mean	2.082 ± 0.272
	Median	1.911	Median	1.961	
	LQ/HQ	1.750 / 2.541	LQ/HQ	1.818 / 2.550	
	Angle (°) Linearity			Angle (°) Linearity	
	Mean	149.808 ± 23.178		Mean	148.819 ± 24.233
	Median	157.290	Median	156.011	
	LQ/HQ	113.331 / 172.988	LQ/HQ	111.911 / 172.680	
Ether OH - FRAGMENT Hits: 2288	Distance (Å)		WATER Hits: 375	Distance (Å)	
	Mean	2.148 ± 0.311		Mean	2.242 ± 0.310
	Median	2.044	Median	2.197	
	LQ/HQ	1.810 / 2.630	LQ/HQ	1.864 / 2.655	
	Angle (°) Linearity			Angle (°) Linearity	
	Mean	145.732 ± 22.467		Mean	141.413 ± 25.116
	Median	150.537	Median	146.358	
	LQ/HQ	112.007 / 172.064	LQ/HQ	105.745 / 171.944	
Nitro OH - FRAGMENT Hits: 258	Distance (Å)		WATER Hits: 48	Distance (Å)	
	Mean	2.286 ± 0.274		Mean	2.327 ± 0.256
	Median	2.295	Median	2.335	
	LQ/HQ	1.916 / 2.653	LQ/HQ	1.920 / 2.683	
	Angle (°) Linearity			Angle (°) Linearity	
	Mean	140.481 ± 21.816		Mean	112.867 ± 23.305
	Median	140.697	Median	113.721	
	LQ/HQ	110.370 / 171.017	LQ/HQ	78.297 / 146.111	

Tertiary amine OH - FRAGMENT Hits: 1152	Distance (Å)		Distance (Å)	
	Mean	2.038 ± 0.335	WATER	Mean
	Median	1.892	Hits: 206	2.128 ± 0.334
	LQ/HQ	1.754 / 2.677		1.970 1.830 / 2.697
	Angle (°) Linearity		Angle (°) Linearity	
	Mean	157.364 ± 19.822		Mean
	Median	164.456		153.871 ± 24.211
	LQ/HQ	125.675 / 174.435		164.250 112.811 / 174.394
Imine OH - FRAGMENT Hits: 1753	Distance (Å)		Distance (Å)	
	Mean	2.043 ± 0.322	WATER	Mean
	Median	1.915	Hits: 451	2.112 ± 0.298
	LQ/HQ	1.741 / 2.624		1.972 1.839 / 2.644
	Angle (°) Linearity		Angle (°) Linearity	
	Mean	158.438 ± 16.609		Mean
	Median	163.094		157.948 ± 17.550
	LQ/HQ	134.840 / 175.023		163.984 130.556 / 175.110
Thioether OH - FRAGMENT Hits: 72	Distance (Å)		Distance (Å)	
	Mean	2.764 ± 0.220	WATER	Mean
	Median	2.871	Hits: 11	2.880 ± 0.150
	LQ/HQ	2.367 / 2.974		2.927 2.416 / 2.974
	Angle (°) Linearity		Angle (°) Linearity	
	Mean	132.033 ± 21.476		Mean
	Median	130.307		107.099 ± 33.218
	LQ/HQ	103.843 / 161.404		99.861 66.977 / 173.008
Thioketone OH - FRAGMENT Hits: 159	Distance (Å)		Distance (Å)	
	Mean	2.404 ± 0.198	WATER	Mean
	Median	2.355	Hits: 59	2.451 ± 0.178
	LQ/HQ	2.207 / 2.702		2.394 2.263 / 2.703
	Angle (°) Linearity		Angle (°) Linearity	
	Mean	156.828 ± 16.969		Mean
	Median	162.246		101.905 ± 13.858
	LQ/HQ	128.712 / 174.013		98.988 86.918 / 122.959
Sulfoxide / Sulfone OH - FRAGMENT Hits: 577	Distance (Å)		Distance (Å)	
	Mean	1.927 ± 0.249	WATER	Mean
	Median	1.861	Hits: 139	1.962 ± 0.225
	LQ/HQ	1.693 / 2.283		1.888 1.776 / 2.333
	Angle (°) Linearity		Angle (°) Linearity	
	Mean	158.131 ± 18.269		Mean
	Median	164.531		158.962 ± 16.307
	LQ/HQ	131.034 / 174.296		164.610 134.046 / 174.127
Phosphine oxide / Phosphone OH - FRAGMENT Hits: 521	Distance (Å)		Distance (Å)	
	Mean	1.724 ± 0.173	WATER	Mean
	Median	1.702	Hits: 133	1.875 ± 0.175
	LQ/HQ	1.540 / 1.898		1.842 1.730 / 2.001
	Angle (°) Linearity		Angle (°) Linearity	
	Mean	165.287 ± 12.754		Mean
	Median	168.385		162.256 ± 15.876
	LQ/HQ	153.613 / 175.640		166.442 145.489 / 174.926

Nitrile OH - FRAGMENT Hits: 240	Distance (Å)		WATER Hits: 30	Distance (Å)	
	Mean	2.020 ± 0.214		Mean	2.092 ± 0.233
	Median	1.962	Median	2.019	
	LO/HQ	1.823 / 2.327	LO/HQ	1.849 / 2.595	
	Angle (°) Linearity			Angle (°) Linearity	
	Mean	157.857 ± 17.242		Mean	160.354 ± 17.386
	Median	163.699	Median	167.028	
	LO/HQ	128.967 / 174.382	LO/HQ	126.289 / 175.973	
Aromatic Ring Acceptor OH - FRAGMENT Hits: 95	Distance (Å)		WATER Hits: 40	Distance (Å)	
	Mean	Lower than 2.6 Å		Mean	2.592 ± 0.150
	Median	2.457 ± 0.098	Median	2.574	
	LO/HQ	2.318 / 2.574	LO/HQ	2.398 / 2.810	
	Angle (°) Linearity			Angle (°) Linearity	
	Mean	146.141 ± 14.614		Mean	86.329 ± 12.634
	Median	146.695	Median	86.645	
	LO/HQ	128.046 / 163.400	LO/HQ	71.237 / 106.280	
Functional groups HBD					
OH - Donors OH - FRAGMENT Hits: 6206	Distance (Å)		WATER Hits: 1266	Distance (Å)	
	Mean	1.925 ± 0.274		Mean	1.796 ± 0.212
	Median	1.828	Median	1.761	
	LO/HQ	1.703 / 2.436	LO/HQ	1.600 / 1.987	
	Angle (°) Linearity			Angle (°) Linearity	
	Mean	156.376 ± 19.611		Mean	162.702 ± 15.046
	Median	162.605	Median	166.697	
	LO/HQ	126.301 / 174.255	LO/HQ	146.216 / 175.585	
NH - Donors OH - FRAGMENT Hits: 1878	Distance (Å)		WATER Hits: 1138	Distance (Å)	
	Mean	2.027 ± 0.233		Mean	1.997 ± 0.233
	Median	1.965	Median	1.942	
	LO/HQ	1.792 / 2.398	LO/HQ	1.752 / 2.335	
	Angle (°) Linearity			Angle (°) Linearity	
	Mean	157.200 ± 15.366		Mean	156.744 ± 16.628
	Median	160.783	Median	160.789	
	LO/HQ	135.371 / 173.138	LO/HQ	134.271 / 173.278	
HPV - CH (Carom) OH - FRAGMENT Hits: 6183	Distance (Å)		WATER Hits: 758	Distance (Å)	
	Mean	2.548 ± 0.121		Mean	2.539 ± 0.133
	Median	2.567	Median	2.563	
	LO/HQ	2.387 / 2.692	LO/HQ	2.354 / 2.693	
	Angle (°) Linearity			Angle (°) Linearity	
	Mean	99.755 ± 24.341		Mean	145.634 ± 16.132
	Median	99.111	Median	145.825	
	LO/HQ	68.091 / 131.628	LO/HQ	124.229 / 167.563	
HPV - CH (Caliph) OH - FRAGMENT Hits: 2985	Distance (Å)		WATER Hits: 305	Distance (Å)	
	Mean	2.588 ± 0.102		Mean	2.588 ± 0.101
	Median	2.611	Median	2.612	
	LO/HQ	2.440 / 2.700	LO/HQ	2.462 / 2.698	
	Angle (°) Linearity			Angle (°) Linearity	
	Mean	94.581 ± 25.792		Mean	147.915 ± 15.037
	Median	92.612	Median	147.765	
	LO/HQ	63.053 / 130.435	LO/HQ	125.438 / 167.102	

Appendix 2: ChemMedChem 2010

Probing Small-Molecule Binding to Cytochrome P450 2D6 and 2C9: An In Silico Protocol for Generating Toxicity Alerts

Gianluca Rossato, Beat Ernst, Martin Smiesko, Morena Spreafico, and Angelo Vedani^{1*}

Drug metabolism, toxicity, and their interaction profiles are major issues in the drug-discovery and lead-optimization processes. The cytochromes P450 (CYPs) 2D6 and 2C9 are enzymes involved in the oxidative metabolism of a majority of marketed drugs. Therefore, the prediction of the binding affinity towards CYP2D6 and CYP2C9 would be beneficial for identifying cytochrome-mediated adverse effects triggered by drugs or chemicals (e.g., toxic reactions, drug–drug, and food–drug interactions). By identifying the binding mode by using pharmacophore prealignment, automated flexible docking, and by quantifying the binding affinity by multidimensional QSAR (mQSAR), we validated a model family of 56 compounds (46 training, 10 test) and 85 compounds (68 training, 17 test) for

CYP2D6 and CYP2C9, respectively. The correlation with the experimental data (cross-validated $r^2=0.811$ for CYP2D6 and 0.687 for CYP2C9) suggests that our approach is suited for predicting the binding affinity of compounds towards CYP2D6 and CYP2C9. The models were challenged by Y-scrambling and by testing an external dataset of binding compounds (15 compounds for CYP2D6 and 40 for CYP2C9). To assess the probability of false-positive predictions, datasets of nonbinders (64 compounds for CYP2D6 and 56 for CYP2C9) were tested by using the same protocol. The two validated mQSAR models were subsequently added to the VirtualToxLab (VTL, <http://www.virtualtoxlab.org>).

Introduction

Human liver cytochromes P450 (CYPs) are a superfamily of oxidoreductase enzymes involved in the metabolism and detoxification of a wide variety of endogenous substrates and xenobiotics. Along with CYP2C18, CYP2C19, CYP2E1, and CYP3A4, the two polymorphic forms of CYP2D6^[1] and CYP2C9^[2] metabolize almost 90% of currently marketed drugs.^[3] Products of these oxidative/metabolic processes may themselves be substrates for conjugating enzymes that catalyze the formation of more hydrophilic and therefore easily excretable metabolites. The metabolic fate of therapeutics and toxic compounds is a fundamental issue in the prediction of pharmacokinetic properties and toxicity estimations during the drug-development process. The capability of CYPs to process structurally diverse compounds is of further interest regarding mechanisms leading to drug–drug and food–drug interactions. Therefore, a predictive in silico model allowing the binding of small molecules towards CYP2D6 and CYP2C9 to be quantified would be highly desirable. The model should also provide information on isoform specificity as well as on the regio- and stereoselectivity of binding.

The X-ray crystal structures of both human CYP2D6^[4] and CYP2C9^[5,6] are available from the Protein Data Bank (<http://www.pdb.org>). Their molecular weights are 54 kDa and their active sites spawn volumes of 797 Å³ (CYP2D6) and 1457 Å³ (CYP2C9).^[7] It is assumed that their low ligand selectivity is associated with the induced-fit abilities of their active sites.^[8] Pharmacophore models have been shown to be valuable tools for predicting metabolic regioselectivity: ligands binding to CYP2D6 often contain a basic nitrogen moiety located some

5–10 Å from the site of metabolism.^[9] In contrast, many ligands binding to CYP2C9 include at least one hydrophobic region between 3–6 Å from a hydrogen-bond acceptor.^[10]

In this work, we emphasize the importance of a correct simulation of the active-site flexibility, comparing atomic mobility from X-ray crystal structure determinations to the average movements within molecular dynamics (MD) simulations. Affinities (*K*) for binding and nonbinding molecules (120 for CYP2D6 and 141 for CYP2C9) were extracted from a dataset of 500 drug-like compounds.^[11] We adopted a mixed-model protocol in which ligand-based prealignment was followed by automated, flexible ligand docking to the protein structure and quantification through multidimensional QSAR (mQSAR).^[12,13] We tested and validated the models by checking the vicinity of the sites of metabolism of the ligands to the iron–heme system, by performing Y-scrambling, and by supplying external datasets of compounds to the models for CYP2D6 and CYP2C9. Compounds displaying affinities $\geq 200 \mu\text{M}$ to the respective two cytochromes were tested by using the same protocol to evaluate the probability of obtaining false-positive predictions. The validated mQSAR models of CYP2C9 and

[a] G. Rossato, Prof. B. Ernst, Dr. M. Smiesko, Dr. M. Spreafico, Prof. A. Vedani
Institute of Molecular Pharmacy
Pharmaceutical Center, University of Basel
Klingelbergstrasse 50, 4056 Basel (Switzerland)
Fax: (+41) 61 267 1552
E-mail: angelo.vedani@unibas.ch

Supporting information for this article is available on the WWW under <http://dx.doi.org/10.1002/cmdc.201000358>.

CYP2D6 were included into the VirtualToxLab (VTL), an *in silico* tool for the prediction of the toxic potential (endocrine and metabolic disruption, interference with the hERG ion channel) of drugs and chemicals.^[14] The VTL currently includes 16 models: 10 nuclear receptors,^[14–22] hERG, and five CYP450^[23] enzymes.

Results and Discussion

First, we correlated the atomic thermal motion from the X-ray crystal structure with the atomic movements observed during MD simulations. Ligand prealignment and automated docking were used to establish a mQSAR study to predict the binding of external compounds. To challenge the models, the consistency with the site of metabolism, sensitivity to biological data (Y-scrambling), together with the prediction of binding and nonbinding compounds were checked. Finally, validated models were inserted in the VTL.

Correlation of experimental and calculated atomic motion

Plasticity, that is, the ability of the active site of a target protein to accommodate and adapt to numerous different substrates,^[24] and protein flexibility are properties shared by several members of the CYP family.^[24,25] These aspects are often utilized to explain their high substrate-binding promiscuity along with high regioselectivity. CYP2D6 and CYP2C9 share more than 40% of their primary sequence and are similar in the fold-

ing and organization of their secondary structural elements.^[4] Still, only 18 out of 56 compounds that bind to CYP2D6 also bind to CYP2C9 with affinities $< 200 \mu\text{M}$ (table 1 in the Supporting Information). To be consistent with the hypothesis of different states of the CYP2C9 active site^[8] and to explain the regioselectivity of binding, a realistic simulation must allow for induced-fit adaptation.^[24] Moreover, the active-site volume calculated for the CYP2D6 structure is too small to allocate some of the known substrates.^[4] To assess the appropriateness of our protein-relaxation protocol, we compared the average displacement (\bar{u}) derived by the thermal displacement factors of the $\text{C}\alpha$ carbon atoms (as obtained from the X-ray crystal structure determination) with the average root-mean-square (rms) value from the MD simulations (Figures 1 and 2).

In CYP2D6 (Figure 1), the $\text{C}\alpha$ carbon atoms of the amino acids in the core of the active site display a rather high stability both in terms of \bar{u} and average rms. Besides the low \bar{u} and average rms of key residues lining the active site (e.g., Phe120, Val370, Glu216, Asp301, and Phe483 are located in a more flexible area (indicated by an asterisk in Figure 1). A larger mobility is observed in the loops B–C and F–G located at the protein surface compared with loops G–H, H–I, and the β sheet β 2–1.

In CYP2C9 (Figure 2), high mobility is displayed by several loops (F–G, G–H, H–I, J–J1), and a β sheet (β 2–1) at the protein surface. As already emphasized,^[8] the active site is rather rigid and the majority of its residues (e.g., Asp293, Thr301) present low values of both \bar{u} and average rms. In comparison, in terms

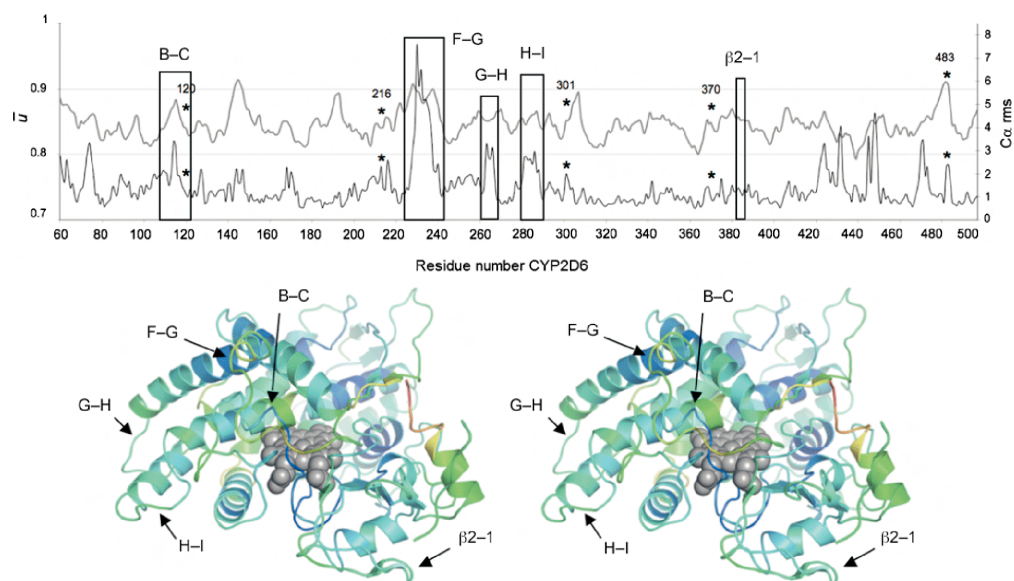


Figure 1. CYP2D6: Average displacement \bar{u} (X-ray; left y axis, light gray, upper curve), and calculated rms (MD; right y axis, dark gray, lower curve) for CYP2D6 after alignment and superposition of 1000 frames recorded along an MD simulation of 1.2 ns for each $\text{C}\alpha$ carbon of each residue (x axis). Different ranges were used between \bar{u} and $\text{C}\alpha$ rms to highlight the similar trends. 3D representation (stereo view) of CYP2D6 with $\text{C}\alpha$ carbons colored by the B factor and the heme portion displayed in a space-filling mode. Loops at the protein surface and the positions of residues important for the binding are highlighted (*).

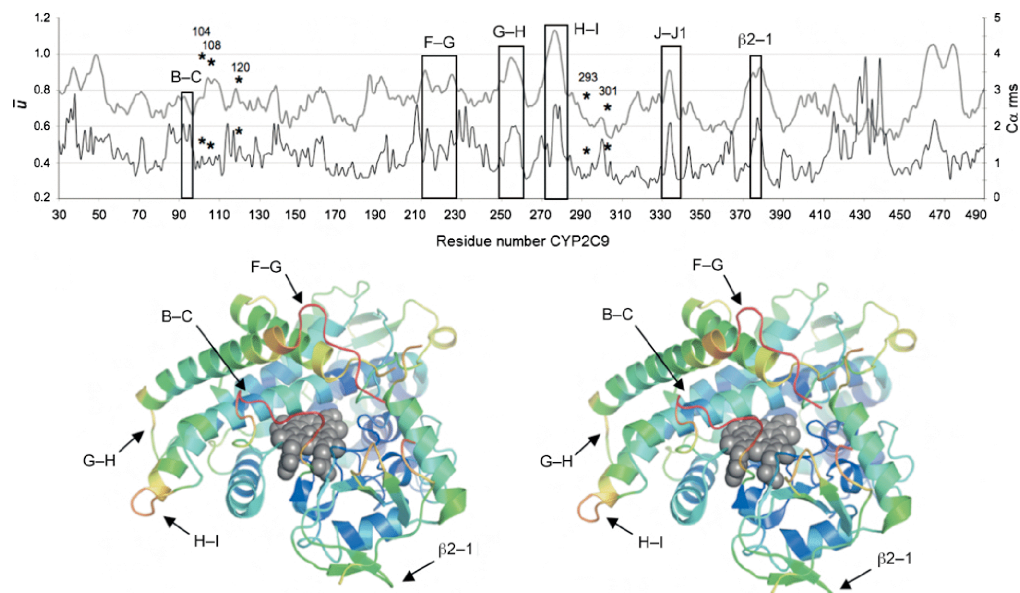


Figure 2. CYP2C9: Average displacement \bar{u} (X-ray; left y axis, light gray, upper curve), and calculated rms (MD; right y axis, dark gray, lower curve) for CYP2C9 after alignment and superposition of 1000 frames recorded along an MD simulation of 1.2 ns for each $C\alpha$ carbon of each residue (x axis). Different ranges were used between \bar{u} and $C\alpha$ rms to highlight the similar trends. 3D representation (stereo view) of CYP2C9 with $C\alpha$ carbons colored by the B factor and the heme portion displayed in a space-filling mode. Loops at the protein surface and the positions of residues important for the binding are highlighted (*).

of \bar{u} , Glu104, Arg108, and Phe120 are more flexible. Finally, the mobility information for Glu104 and Arg108 largely differs between \bar{u} and calculated rms.

This discrepancy between the MD simulations and X-ray structural information could arise from an incorrect handling of the system by the force field. Furthermore, the length of the simulation (1.2 ns), simulation conditions, and water model used to reproduce the X-ray determination conditions could also contribute to the observed discrepancy. Of course, in the case of the mutated residues that were reset to the wild type (e.g., D230L in CYP2D6, Figure 1), some inconsistency is to be expected. The aforementioned comparison only offers qualitative information on enzyme motion owing to numerous limiting factors such as atom selection (i.e., only $C\alpha$ carbon atoms), atom position (i.e., atoms belonging to loops and at the enzyme surface present higher fluctuations), and difficulties in comparing absolute values for single atoms out of the two experiments (X-ray and MD simulation). Nonetheless, the numerous agreements allow a general interpretation of enzyme motion in a true biological system.

Prealignment and automated, flexible docking

The correlation of \bar{u} and average rms (Figure 1 and 2) further underlines the role of protein flexibility and the requirement of a flexible-docking protocol that simulates the binding to CYP2D6 and CYP2C9. Previous studies^[9,10,26] provided evidence

such that, despite the high promiscuity in ligand binding of CYPs, common pharmacophore features can be identified for the different isoforms. To include this binding information in our approach, we have adopted a pharmacophore-based prealignment, by using methdilazine as template for CYP2D6 (Figure 3) and trifluoperazine, medrysone, piroxicam, and F02 for CYP2C9 (Figure 4). Because we identified only one chemical class (i.e., tricyclics) for CYP2D6, a unique template was used for the dataset. For CYP2C9 the compounds displaying the strongest binding affinity among the three chemical classes identified (i.e., tricyclics, steroids, sulfonamides) were used as templates. Piroxicam and F02 refer to the sulfonamide-like ensemble but were used separately because they display distinct pharmacophore features.

Ligand prealignment was followed by a combinatorial scan of amino acid side chains in the active site to identify possible hydrogen-bond partners and to rearrange the active site consequently. This active-site configuration represented the starting point for our flexible-docking protocol as implemented in Cheetah.^[17,23,27] In this approach, the protein backbone is kept rigid but the amino acid side chains are allowed to move. The docking output based on the prealigned conformations was energetically scored towards docking poses obtained through an extended Monte Carlo search. For 75% of the CYP2D6 and 61% of the CYP2C9 ligands, at least one structure obtained from the prealignment by Alignator^[28] was retained in the ensemble of the best-scored poses, which justifies the use of

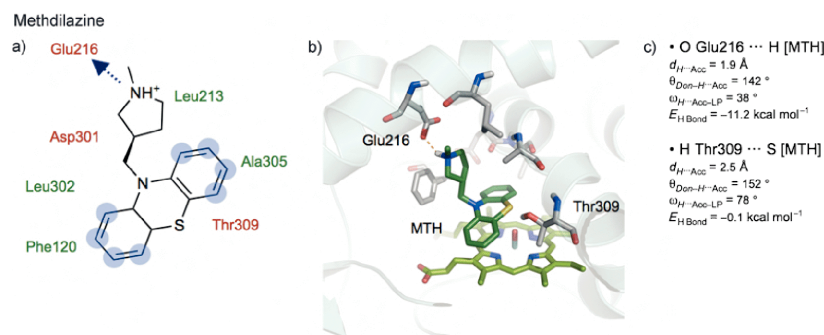


Figure 3. CYP2D6: Details of binding for the template methdilazine used for the prealignment. a) 2D representation of the active site is shown (polar amino acids are depicted in red; hydrophobic amino acids in green; atoms of the ligand exposed and not interacting with any residue of the enzyme are blurred in blue). b) 3D representation of the binding mode is given (residues of the binding site are shown as gray; methdilazine as dark green; and the heme as light-green sticks). c) Details of the main interactions in terms of distance (d), angle (θ), lone-pair deviation (ω), and energy contribution (E_{HBond} [kcal mol⁻¹]) between methdilazine and CYP2D6 are listed.

both the prealignment algorithm and the Monte Carlo searching. A flowchart describing the prealignment and docking protocol is provided in Figure 5.

In the binding mode, methdilazine docked to CYP2D6 (Figure 3) is accommodated in a hydrophobic pocket lined by Leu213, Leu302, Phe120, and Ala305. Its tricyclic moiety is located in close vicinity to the heme, possibly interacting through π - π interactions. The positively charged nitrogen atom that forms a strong electrostatic interaction with Glu216 (details of the interactions in Figure 3), together with the aromatic ring located close to the heme, correctly resemble the pharmacophore hypothesis for CYP2D6.^[10] Accessibility of the aromatic carbons of the ligand (highlighted in Figure 3) by the iron atom of the heme agrees with the experimental data,^[29] which identify these atoms of tricyclic compounds as privileged sites of metabolism.

Previous site-directed mutagenesis experiments underline the importance of residues Phe120, Glu216, and Asp301 for binding.^[30,31] All of them are involved in the binding mode, especially Glu216, which interacts with ligands containing a positively charged nitrogen at physiological pH (47 out of 56 ligands). In principle, the negatively charged residue Asp301 could also face the active site and interact with the positive nitrogen from diverse ligands. However, a structural role in protein folding and heme incorporation^[32] is attributed to this residue. Finally, Phe120 was shown to influence the regioselectivity of substrate oxidation (e.g., the mutant F120L increases K_m for dextromethorphan by a factor of 30).^[31]

The four templates (trifluoperazine, medrysone, piroxicam, and F02; Figure 4) bind to a similar area of the active site of CYP2C9. The hydrophobic residues (Val113, Leu208, Ala297) line the channel leading to the heme and contribute hydrophobic interactions in the different binding modes of the four templates. In the trifluoperazine complex with CYP2C9 (Figure 4 a), the electrostatic interaction between the positively charged nitrogen of the ligand and Asp293 acts as an

anchor, which orients the aromatic rings of the ligand orthogonal to the side chain of Phe476.

The binding mode proposed for medrysone (Figure 4b) locates the rigid chemical scaffold further away from the heme system. Two hydrogen bonds are established between the ligand and protein: 1) between the carbonyl of ring A and the backbone NH of Leu102 and 2) between the oxygen atom of the hydroxy group of the ligand and the side chain of Arg108.

Although piroxicam and F02 (Figure 4c, d) share a sulfonamide pharmacophore, they have been selected for additional functional groups: amide and pyridinium moieties in piroxicam, and a sulfur atom and a second sulfonamide in F02. Piroxicam establishes a hydrogen bond between the sulfonamide and the side chain of Arg108, and an electrostatic interaction between the pyridinium moiety and the side chain carboxylate of Asp293. One sulfonamide of F02 similarly interacts with the side chain of Arg108 allowing the aromatic ring, located at the edge of ligand side chain, to establish a close contact with the iron atom of the heme.

Similar to the finding for CYP2D6, the privileged site of metabolism by CYP2C9 of the ligands trifluoperazine,^[33] medrysone,^[34] and piroxicam^[35] are considered to be exposed to and accessible for the iron in the proposed binding modes (Figure 4a–d). No data have been found for F02. Nonetheless, owing to the chemical similarity with piroxicam, we can hypothesize that the site of metabolism is located on the aromatic ring docked in a position accessible to the heme.

Multidimensional QSAR

Up to four poses within an energy window of 10 kcal mol⁻¹ were selected for each of the 56 ligands of CYP2D6 and 85 ligands of CYP2C9, including different stereoisomers and protonation states for a total of 216 poses for CYP2D6 and 351 poses for CYP2C9. The 4D ensemble of the docked poses was used as the input for Quasar.^[12,13]

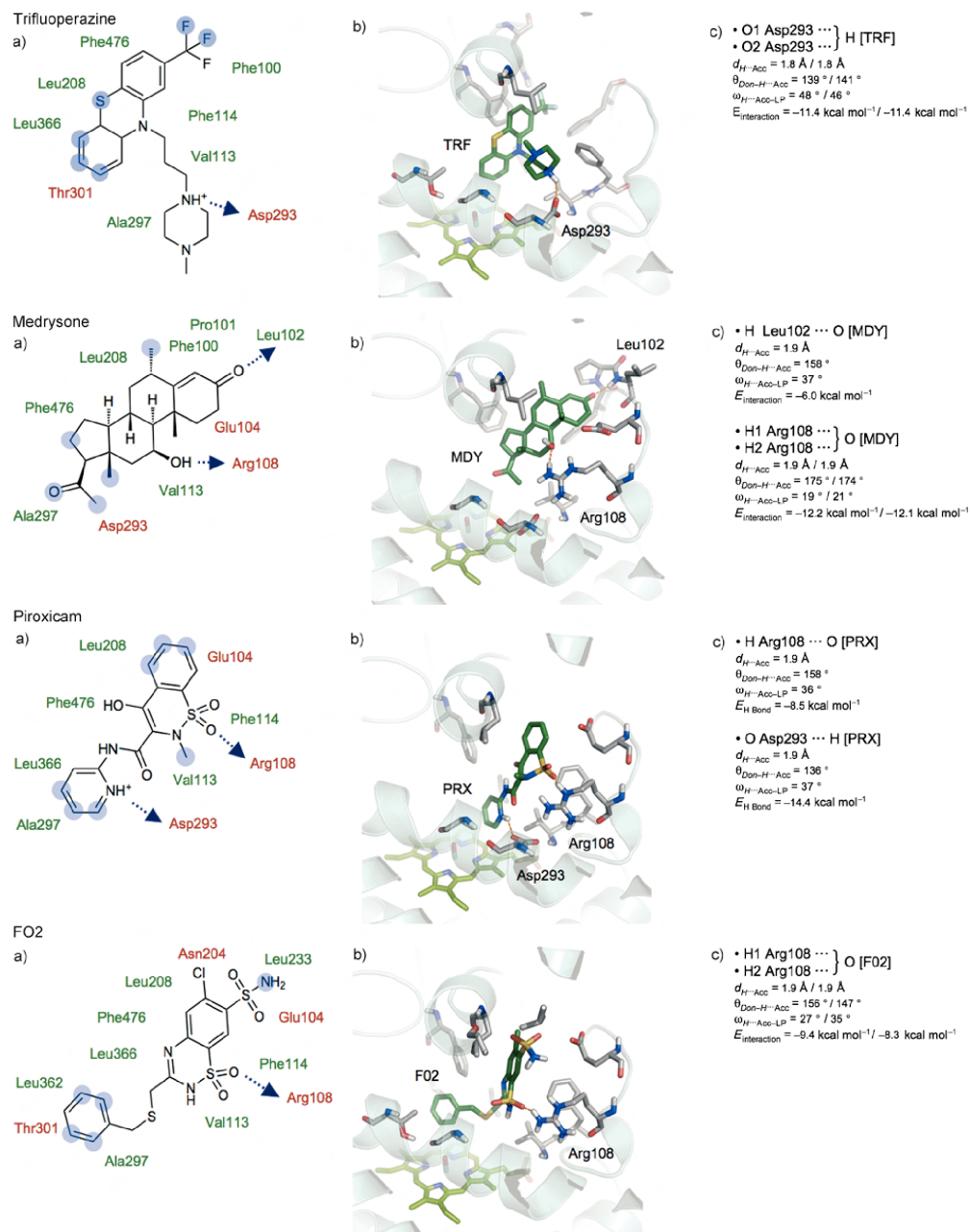


Figure 4. CYP2C9: Binding details of the prealignment templates trifluoperazine, medrysone, piroxicam and F02. In each case: a) 2D representation of the active site is shown (polar amino acids are depicted in red; hydrophobic amino acids in green; atoms of the ligand exposed and not interacting with any residue of the enzyme are blurred in blue). b) 3D representation of the binding mode is given (residues of the binding site are shown as gray; ligand as dark green; and the heme as light-green sticks). c) Details of the main interactions in terms of distance (d), angle (θ), lone-pair deviation (ω), and energy contribution ($E_{H bond}$ [kcal mol⁻¹]) between the ligand and CYP2C9 are listed.

The model family of CYP2D6 (Figure 6a and Table 1) converged at a cross-validated r^2 value of 0.811 for the 46 training compounds by leaving one third of the set out, and yielded a predictive r^2 of 0.711 for the ten test ligands. The average deviation (rms) between experimental and calculated affinities was of a factor of 1.2 for the training and 2.4 for the test set. The maximal deviation in the prediction of binding affinities for a compound was of a factor of 5.0 for the training set and 8.5 for the test set.

The model family of CYP2C9 (Figure 6b and Table 1) converged at a cross-validated r^2 of 0.687 for the 68 training compounds by leaving one fifth of the set out. The predictive r^2 value for the 17 test ligands was 0.423. The p^2 value was only moderate owing to the distribution of the affinities for CYP2C9 together with the rather narrow range of affinities (3.25 orders of magnitude). The calculation of the regression coefficient (q^2) was mainly influenced by the affinities distribution and from the corresponding deviation from the average experimental affinity value as in [Eq. (1)]. The predictive r^2 (test set) was also influenced by the distribution and range of affinities of the test set compounds [Eq. (1)]. Thus, the relatively modest q^2 value was not a consequence of poor prediction but of the limited data available.

$$q^2 = 1.0 - \frac{\sum(\Delta G_{\text{pred}} - \Delta G_{\text{exptl}})^2}{\sum(\Delta G_{\text{exptl}} - \Delta G_{\text{exptl average}})^2} \quad (1)$$

This unavoidable inhomogeneity of distribution is visible in Figure 6b in which the majority of ligands bind in the low-affinity range. This observation further correlates with the promiscuous character of these isoforms of enzymes. Despite the mentioned values, the worst predicted compound deviates from the experimental value by only a factor of 4.4 for the training set and of a factor of 4.8 for the test set. The average rms consistently reflects a low deviation between experimental and predicted binding affinity both for the training and the test set, diverging by a factor of 1.2 and 1.7, respectively.

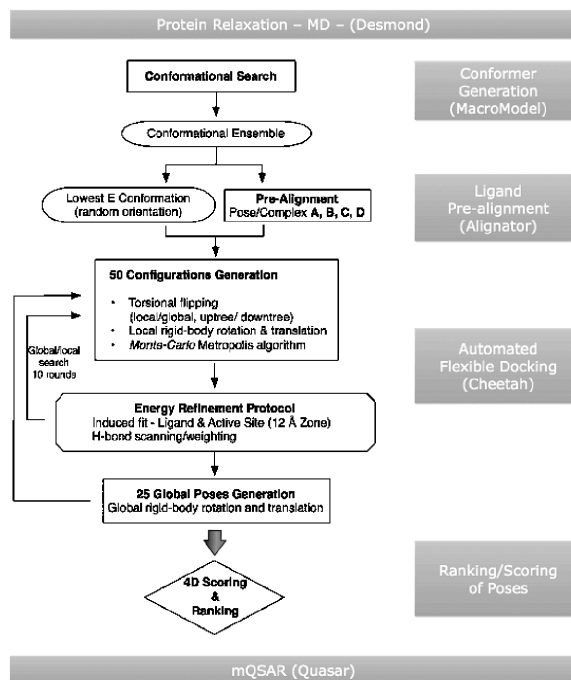


Figure 5. Protein relaxation, prealignment, and flexible-docking protocol. For each ligand, a conformational search generates a pool of conformers of which those within 5 kcal mol⁻¹ from the lowest-energy conformation found were considered. Each conformer of the ensemble is prealigned to a template, thereby allowing side-chain rearrangement within the active site (software Alignator). Up to four of the different ligand-protein complexes (complexes A–D) generated were used as starting poses in the automated, flexible docking (software Cheetah). All configurations obtained by either the minimized prealigned structures or the Monte Carlo sampling were scored and ranked (additional information is provided in table 4 in the Supporting Information).

Consistency with the sites of metabolism

Attempting to evaluate the reliability of the docking protocol and the subsequent quantification by mQSAR, we visually inspected the best-ranked poses for some of the ligands in the dataset. The aim was to check the accessibility of the experimentally known sites of metabolism for the iron atom of the heme during a hypothetical metabolic hydroxylation process.

In Figure 7a–c, the vicinity of the sites of metabolism to the heme for three compounds binding to CYP2D6 are shown. CYP2D6 catalyzes the transformation of nortriptyline (Figure 7a) to its metabolite 10-hydroxynortriptyline through a hydroxylation process.^[36] Within the docking selection (see the selection criteria above) the carbon atom in position 10 is accessible to the heme; similarly the sites of metabolism for the other two tricyclic antidepressants (clomipramine and imipramine), propranolol, and ondansetron^[26] are matched and are accessible in the docking poses obtained from this protocol. O-Demethylation is another metabolic reaction processed by CYP2D6. Galantamine (Figure 7b) undergoes this process and

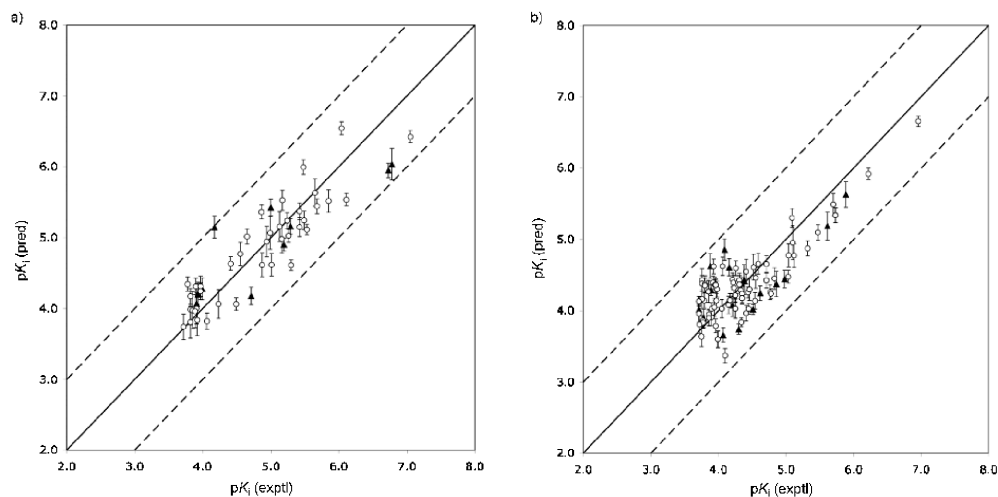


Figure 6. Comparison of experimental (exptl) and predicted (pred) binding affinities of the training set (\circ) and test set (\blacktriangle) for a) CYP2D6 and b) CYP2C9. Dashed lines are drawn at a factor 10 from the experimental values.

	r^2	q^2	rms training	max rms training	p^2	rms test	max rms test
CYP2D6	0.815	0.811	1.2	5.0	0.711	2.4	8.5
CYP2C9	0.695	0.687	1.2	4.4	0.423	1.7	4.8

[a] Rms training and rms test: average rms deviation from the experimental value for training and test set, respectively; max rms training and max rms test: maximum deviation from the experimental value noticed for training and test set, respectively; q^2 : cross-validated r^2 ; p^2 : predictive r^2 .

its methoxy group is metabolized by oxidative dealkylation into the corresponding hydroxy group.^[37] In the docking pose the vicinity of this functional group to the heme is consistently found. Clomipramine (Figure 7c), in a similar way to nortriptyline, can be hydroxylated in position 8,^[38] which is accessible to heme in the docking pose obtained.

The sites of metabolism of tolbutamide, dapsone, losartan, chlorpropamide, diclofenac, and phenytoin^[39] are linearly accessible for the heme of CYP2C9. However, the distances from the oxygen atom of the iron–oxygen system is rather high (e.g., dapsone 4.2, diclofenac 5.2, phenytoin 7.6 Å). One reason could be the large volume of the active site. In addition, in the X-ray structure of CYP2C9 in complex with warfarin,^[6] the ligand is located in a distal part of the active site that is 10 Å away from the iron of the heme. Williams et al.^[6] suggested that this secondary pocket could represent a primary recognition site or a binding pocket for competitive inhibition of the enzyme. In Figure 7d–f examples of the docking poses and

their relative position to the heme for dapsone, diclofenac, and phenytoin are shown.

Y-scrambling and binding of external datasets to CYP2D6 and CYP2C9

To investigate the sensitivity and to scan the predictive capability both towards binding and nonbinding compounds further, the two models were subject to Y-scrambling. An ensemble of 79 compounds for CYP2D6 (15 external binders and 64 nonbinders) and of 96 compounds for CYP2C9 (40 external binders and 56 nonbinders) were tested by the respective models to check the relative deviance from the known experimental binding affinities. The Y-scrambling technique^[40] shuffles the experimental binding-affinity data on which the model is built. If the prediction of the test set reproduces a good correlation by using this “random” model, this is an indication that the sensitivity towards the biological data is poor. Forty Y-scrambling tests (that used a different random-number seed) were performed for each target, which utilized the same settings used for the “normal” models. Out of the 40 models, the average predictive r^2 was of -0.318 for CYP2D6 and -0.283 for CYP2C9 (tables 5 and 6 in the Supporting Information), demonstrating the sensitivity of our models towards the biological data. When QSAR models are established, it is important to avoid the overfitting of these models. To handle this problem correctly, the evolution of predictive r^2 and cross-validated r^2 were monitored every ten generations, to interrupt the evolution of the genetic algorithm subsequently when the predictive r^2 was starting to drop while the cross-validated r^2 was continuing to rise.

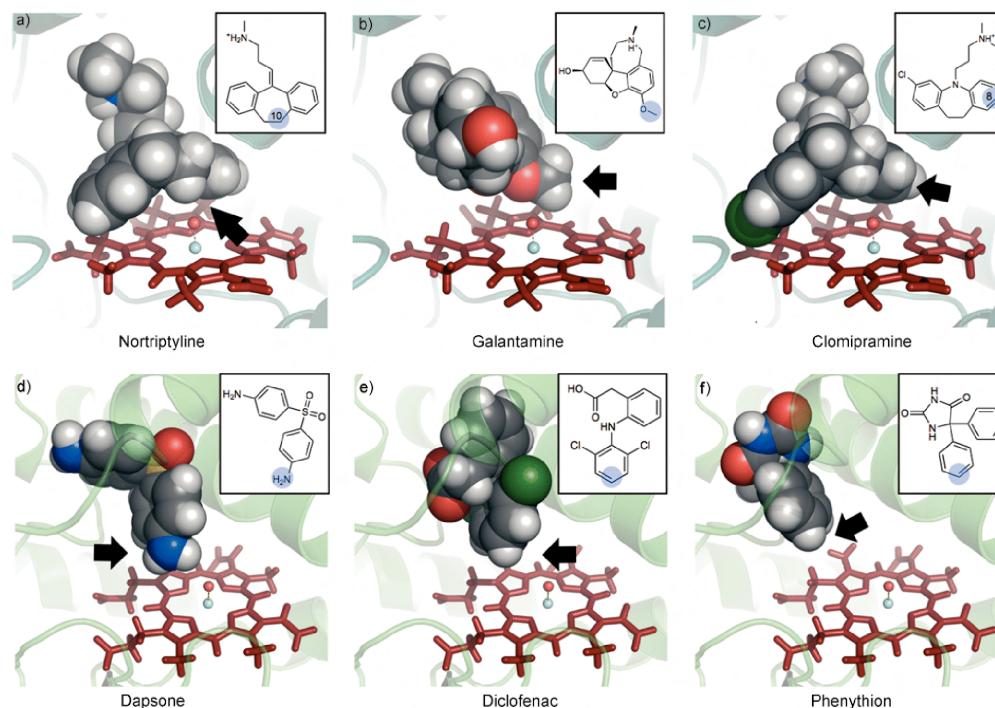


Figure 7. Docking poses as obtained from prealignment and automated, flexible docking: a–c) Nortriptyline, galantamine, and clomipramine, respectively, binding to CYP2D6; d–f) dapsone, diclofenac, and phenythion, respectively, binding to CYP2C9. The secondary structure of the enzyme is represented by cartoons, ligands are in the space-filling mode, the heme portion is represented by red sticks, and iron and oxygen are represented as spheres. Black arrows indicate the primary sites of metabolism as obtained from the experimental data.

In addition, we have simulated the binding of six classic anti-histamines^[41] and nine compounds from the Strobl group^[42] within the same binding range as the training set. In this simulation 60% of the ligands have been predicted within a factor of 10 in K_i from the experimental data, 86.7% within a factor of 25, and 13.3% with more than a factor of 25 (Table 2). In a similar way, we have simulated the binding of 40 different inhibitors to CYP2C9^[43] within the same range of binding affinities used to build up the model. In this instance, 80% of the ligands have been predicted within a factor of 10 in K_i from the experimental data, 87.5% within a factor of 25, and 12.5% differ more than a factor of 25 (Table 3).

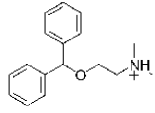
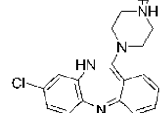
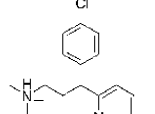
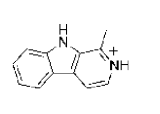
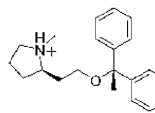
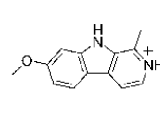
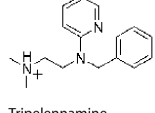
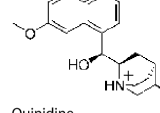
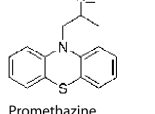
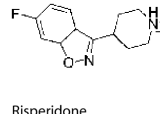
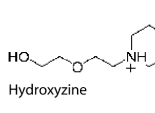
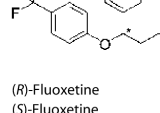
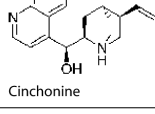
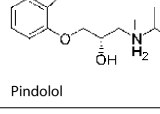
The data used to build the model originate from a filtered Merck repository^[11] of drug-like compounds in which the majority of the compounds present affinities either to CYP2D6 or to CYP2C9 that are equal to or higher than 200 μM . To challenge our model further and to evaluate the probability of obtaining “false-positive” affinity values for low-affinity or non-binding compounds, we simulated the binding of 64 nonbinders to CYP2D6 and 56 of nonbinders to CYP2C9. The simulations were performed by using the protocol implement in VTL,^[14] in which the ligands pass through an extensive (5000 steps) conformational search in aqueous solution, protonation,

and tautomeric-state calculation at physiological pH followed by prealignment of the low-energy conformation of the ligands, flexible docking, and binding-affinity prediction by using mQSAR.^[12,13] For CYP2D6, 56 out of 64 (85.9%), and for CYP2C9, 46 out of 56 (82.1%) ligands were predicted to be nonbinders or to bind within a factor of 10 in K_i from the weakest binder of the training set (praziquantel for 2D6 and protriptyline for 2C9; Figure 8).

VTL dataset screening

The models presented herein for CYP2D6 and CYP2C9 were added to the VTL,^[14] an in silico tool for predicting the toxic potential (endocrine and metabolic disruption) of drugs, chemicals, and natural products. Through automated flexible docking combined with mQSAR it simulates and quantifies the interactions of drug-like compounds to 16 different proteins (the androgen,^[15] aryl hydrocarbon,^[16] estrogen α/β ,^[17] glucocorticoid,^[18] liver X,^[19] mineralocorticoid,^[20] peroxisome proliferator-activated receptor γ ,^[21] and thyroid α/β receptors;^[22] the enzymes CYP1A2, CYP2A13, CYP2C9, CYP2D6, CYP3A4,^[23] and the hERG ion channel^[44]).

Table 2. External dataset of compounds that bind to CYP2D6 giving the compound structure and name along with experimental (exptl) and calculated (calcd) binding affinities.

Compound	Exptl K_i [μM]	Calcd K_i [μM]	Compound	Exptl K_i [μM]	Calcd K_i [μM]
 Diphenhydramine	11.7	28.8	 Clozapine	19.0	33.2
 Chlorpheniramine	10.8	25.9	 Harmine	85.6	960.2
 Clemastine	2.0	44.4	 Harmine	50.0	522.8
 Tripeleminamine	5.6	81.5	 Quinidine	4.6	156.4
 Promethazine	3.8	11.8	 Risperidone	21.4	5.3
 Hydroxyzine	3.9	1.6	 (R)-Fluoxetine (S)-Fluoxetine	1.4 0.2	3.1 178.5
 Cinchonine	3.5	71.9	 Pindolol	4.8	11.2

The VTL repository of 2518 drugs, chemicals, and natural products have been screened and the predicted toxicity alerts are listed at the website: <http://www.virtualtoxlab.org>. It should be noted that amphetamine and methylenedioxyamphetamine are predicted to bind to CYP2D6 with a K_i of 8.9 and 1.6 μM , respectively. In the literature, both methamphetamine and methylenedioxymethamphetamine are metabolized by CYP2D6.^[45] Similarly, sertraline displays potential for media-

ting drug–drug interactions as well as inhibition of CYP2D6.^[46] Our totally automated protocol predict sertraline as a high-affinity binder to CYP2D6 with a K_i of 4.1 μM . A recent survey reported pantoprazole as one of the most potent CYP2C9 in vitro inhibitors among the currently used proton pump inhibitors (PPI) with a K_i value of 6.5 μM .^[47] The screening of the VTL repository on CYP2C9 scored pantoprazole as one of the most affine binders to CYP2C9 with a predicted K_i of 33.1 μM , indi-

Table 3. External dataset of compounds that bind to CYP2C9 giving the compound structure and name along with experimental (exptl) and calculated (calcd) binding affinity.

Compound	Exptl K_i [μM]	Calcd K_i [μM]	Compound	Exptl K_i [μM]	Calcd K_i [μM]		
	E1-(9)R E1-(9)S	31.4 5.9	7.0 4.0		E14	15.0 25.8	
	E-(9)R(11)R E2-(9)S(11)S E2-(9)R(11)S E2-(9)S(11)R	7.0 12.0 20 10.0	36.0 88.5 79.3 155.3		E15	50.0 193.0	
	E3	0.6	6.9		E16	2.0 181.3	
	E4-(9)R E4-(9)S	0.6 0.5	10.5 39.9		E17	0.2 10.3	
	E5-(9)R(11)R E5-(9)S(11)S E5-(9)R(11)S E5-(9)S(11)R	37.0 1.0 14 0.1	11.2 3.4 1.3 31.3		E18	18.0 6.8	
	E6-(9)R E6-(9)S	10.0 11.0	12.0 15.0		E19	48.0 8.9	
	E7-(9)R E7-(9)S	29.0 21.0	10.3 20.5		E20	25.0 14.6	
	E8-(9)R E8-(9)S	13 17	87.6 22.0		E21	7 25.1	
	E9-(9)R E9-(9)S	14 8	59.7 15.7		E22	1.5 12.5	
	E10	10	26.0		E23	1.1 5.5	
	E11	70	28.9		E24	18 46.3	
	E12	11	11.3		E25	3 26.2	
	E13	45	94.8		E28 E29 E30 E31	2.4 2 6 0.1	0.3 8.7 5.5 7.7

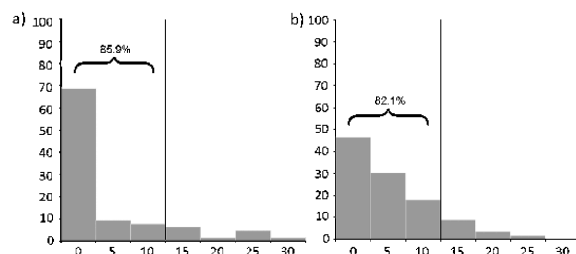


Figure 8. Plotted percentages of compounds (y axis) and error factor in K_i (x axis) from the weakest binder of a) CYP2D6 and b) CYP2C9. The vertical line isolates the compounds predicted to have a K_i within a factor of 10 from the weakest binder of the two datasets.

ating an appreciably small difference between experimental and predicted affinity (a factor of five).

Conclusions

The aim of this study was to establish a reliable approach for the prediction of drug-like molecules binding to CYP2C9 and CYP2D6. The promiscuity of these enzymes is often attributed to their flexibility and their large active-site volume. Comparison of the mobility of $C\alpha$ carbon atoms during the X-ray crystal structure determination and MD simulations confirmed the high flexibility of the external loops. On the contrary, the $C\alpha$ carbon atoms of the active site are relatively rigid. To simulate the binding to CYP2D6 and CYP2C9 correctly, we accounted for side-chain flexibility during the prealignment and docking protocol. The alignment of a 4D ensemble of 56 compounds for CYP2D6 and 85 for CYP2C9 was used to generate mQSAR models (software Quasar). This 3D model represents the active site as a surrogate characterized by quasi-atomistic properties reflecting those of the true biological receptor.

The models yielded a cross-validated r^2 of 0.811 for CYP2D6 and 0.695 for CYP2C9. The lower value of predictive r^2 for CYP2C9 (i.e., 0.423) compared with CYP2D6 (i.e., 0.711) was due to the high concentration of compounds displaying low binding affinities as well as to the narrow range of affinities. Nonetheless, the worst predicted compound in the cross-validation procedure was within a factor of 5.0 for CYP2D6 and 4.4 for CYP2C9. Model sensitivity was assessed by 40 scramble tests for each of the two isoforms, which yielded an average predictive r^2 of -0.318 for CYP2D6 and -0.283 for CYP2C9. Visual inspection of the docked poses displayed consistency between the location of sites of metabolism obtained from experimental studies and the linear accessibility of these sites to the heme in the docked complexes. When models were tested on an external dataset, the experimental values were predicted within a factor of 10 for 60% of the compounds for CYP2D6 and for 80% of the compounds for CYP2C9 (86.7 and 87.5% within a factor of 25 for CYP2D6 and CYP2C9, respectively). The binding properties for 64 nonbinders or weak-binders for CYP2D6 and 56 for CYP2C9 allowed us to test the probability of obtaining false-positive alerts from our models. In this ac-

count, 85.9 and 82.1% were predicted to be non-binders or within a factor of 10 from the weakest binding compound from the training set.

The two CYPs models have been added to the VTL developed by the Biographics Laboratory 3R for the in silico prediction of metabolic and endocrine disruption and interference with the hERG ion channel. Presently, it includes 16 validated models for five CYPs, 10 nuclear receptors, and the hERG ion channel.

Our study underlines the fact that for the correct modeling of drug-like molecules binding to CYP2D6 and CYP2C9, ligand-pharmacophore features, state of oxidation of the heme, and active-site motion must be considered. Previously published models are either based on pharmacophore hypotheses^[9,48,49] or physico-chemical and structural descriptors.^[50,51] Because the binding affinity of small-molecule ligands depends highly on the protein–ligand interactions, and because the experimental structures of both CYP2D6 and CYP2C9 are available, it is advantageous to identify the binding mode through flexible docking at these entities. In contrast to other approaches in the field, this allows the simulation of the induced fit. The mQSAR technology used for estimating binding affinities would seem to be necessary because it allows protein–ligand interactions to be quantified at the 4D-structural level, as well as the induced fit (5D) and solvation phenomena (6D). Our results suggest that mixed-model QSAR^[18] that includes ligand prealignment is not only a reliable tool for predicting sites of metabolism and particularly for estimating associated binding affinities, but also for producing toxicity alerts.

Experimental Section

Protein structures: For CYP2D6, we employed the only available 3D structure of the human apoenzyme (PDB code: 2F9Q, 3.0 Å resolution, $R_{\text{free}} = 0.286$).^[4] For our simulations, the mutations that were used to facilitate crystallization were reset to those of the wild type, that is, D230 back to L230, R231 back to L231, and M374 back to V374.

For CYP2C9, we selected the crystal structure in complex with flurbiprofen (PDB code: 1R9O, 2.0 Å resolution, $R_{\text{free}} = 0.236$).^[5] The residues (V38-I39-G40-N41-I42 and Q214-I215-C216-C217-N218-F219-S220) that were not resolved in the electron-density map of the X-ray structures were rebuilt by using the PrGen software.^[52]

Prior to the MD sampling with Desmond,^[53] we applied a relaxation protocol divided into 1) molecular-mechanics minimization with restrained and unrestrained solute, 2) two MD simulations of 12 ps at a temperature of 10 K in the NVT (in the NVT or canonical ensemble, during the MD simulation the parameters of moles (N), volume (V), and temperature (T) are considered to be conserved) and NPT ensemble (in the NPT or isothermal–isobaric ensemble, moles (N), pressure (P), and temperature (T) are considered to be conserved) with non-hydrogen solute atoms restrained, 3) two simulations of 24 ps in the NPT ensemble at 300 K, 1 atm with restrained and unrestrained water oxygen atoms. Two MD simulations of 1.2 ns were then performed by using CYP2D6 in complex with methildazine and CYP2C9 with the (co-crystallized) ligand flurbiprofen.^[5]

The enzymes were simulated within cubic boxes with a volume of around $60 \times 100 \times 100 \text{ \AA}$ in size by using the TIP3P water-solvation model [in the TIP3P water model, water molecules are represented through a three-interactions-site model (i.e., for each atom of the molecule is assigned a point charge)], periodic-boundary conditions, and a salt concentration (NaCl) of 0.15 M, mimicking physiological conditions.

To analyze the mobility of the C α atoms, we correlated the thermal displacement factor (B_{isor} , which is the B factor or thermal displacement factor). It reflects the mean displacement of an atom $\langle u \rangle$, thus the magnitude of oscillation), as obtained from the X-ray structure determination, with motions observed in our MD simulations (rms deviations analyzed at 1.2 ps intervals). Alignment, fit, and rms deviations were calculated for each frame of the simulations relative to the starting frame of the trajectory by using the McLachlan algorithm^[54] as implemented in the program ProFit (A. C. R. Martin, <http://www.bioinf.org.uk/software/profit>). Next, a trajectory-cluster analysis by using the hierarchical cluster linkage method was used to identify the most populated clusters of structures. From these clusters, the structure displaying the lowest energy was fully refined in aqueous solution by using the OPLS 2005 force field,^[55] as implemented in the software MacroModel.^[56] The atomic partial-charge model for all ligands (CM1) was generated with AMSOL,^[57] partial charges for the proteins were assigned from the AMBER database.^[58]

The iron–heme system was modeled in the oxenoid iron (formally $[\text{Fe}^3 + \dots \text{O}^{2-}]$) as in a previous study for CYP3A4.^[23] To allow for a realistic simulation, we used a directional force field (Figure 9), which also features a metal function allowing for charge transfer (particularly among oxygen, iron, sulfur, and the heme nitrogen atoms) and for functions controlling symmetry, directionality, and ligand-field stabilization (LFSE).^[59] In particular, the dynamic charge-transfer function enables for a smooth and continuous transition between ionic and covalent character of the metal–ligand bond.

$$E_{\text{total}} = \sum_{\text{bonds}} K_r (r - r_{\text{eq}})^2 + \sum_{\text{angles}} K_\theta (\theta - \theta_{\text{eq}})^2 + \sum_{\text{torsions}} \frac{V_n}{2} [1 + \cos(n\phi - \gamma)] + \sum_{\text{nb pairs}} \frac{q_i q_j}{4\pi\epsilon_0 D(r) r_{ij}^2} + \sum_{\text{nb pairs}} \left(\frac{A}{r_{ij}^{12}} - \frac{B}{r_{ij}^6} \right) + \sum_{\text{H bonds}} \left(\frac{C}{r_{ij}^{12}} - \frac{D}{r_{ij}^6} \right) \cdot \cos^2(\theta_{\text{Don-H}\dots\text{Acc}}) \cdot \cos^6(\omega_{\text{H}\dots\text{Acc-LP}}) + \sum_{\text{metal pairs}} \frac{q_i^{\text{CT}} q_j^{\text{CT}}}{4\pi\epsilon_0 D(r) r_{ij}^2} + \sum_{\text{metal pairs}} \left(\frac{E}{r_{ij}^{12}} - \frac{F}{r_{ij}^6} \right) + (E_{\text{MC}} + E_{\text{LFS}}) \cdot \prod_{\text{angles}} \cos^2(\Psi_{\text{Lig-Met-Lig}} - \Psi_{\text{eq}}) \cdot \frac{1}{n} \sum_{\text{ligands}} \cos^n(\omega_{\text{Met}\dots\text{Lig-LP}})$$

Figure 9. Yeti force field^[59] with metal-function terms highlighted.

Ligand structures and affinity data: The experimental affinities of 120 compounds (56 binders and 64 nonbinders) for CYP2D6 and 132 compounds (76 binders and 56 nonbinders) for CYP2C9, were kindly provided by D. R. McMasters (structures of compounds in tables 1 and 2 in the Supporting Information).^[11] Compounds that exhibit poor solubility, non-Michaelis–Menten kinetics, and display fluorescent interference during the competitive inhibition assay used to obtain their affinities were excluded. The selection of inhibition data originates from a unique dataset of 500 drug-like compounds from the Merck sample repository, measured by using recombinant enzymes. We selected a subset according to the fol-

lowing criteria: $K_i < 200 \mu\text{M}$, compounds associated with a trade name (to identify the active stereoisomer), and ligands for which the database identifier or smiles code were unambiguous. Whereas for CYP2D6 the activity range of the employed compounds extended over 3.3 orders of magnitude, the range of CYP2C9 was only 2.4. Therefore, we added nine compounds from a different source^[60] (table 3 in the Supporting Information), yielding a total of 141 ligands for the CYP2C9 data set. The selection used to estimate the potential of false-positive predictions (64 for CYP2D6 and 56 for CYP2C9) originates from compounds with affinities $\geq 200 \mu\text{M}$ of the released Merck sample repository. The most probable protonation state at pH 7.4 was predicted through the software Epik from the Schrodinger software suite,^[56] and in case of proton dissociation constants close to pH 7.4, both protonation states were examined.

After generation, the 3D structures of all ligands were subjected to a conformational search in aqueous solution (5000 rounds of Monte Carlo search) performed with MacroModel from Schrodinger Inc.^[56] For each ligand, all conformations within 10 kcal mol^{-1} from the lowest-energy entity were retained and composed into a 4D dataset by taking in account different conformations, poses, protonation states, and tautomers.

Prealignment and automated, flexible docking: Because the crystal structures of both cytochromes^[4–6] do not contain any of the dataset ligands, the template complexes were generated by manual docking. The template molecules were the most affine ligands, identified for the different chemical classes, which were manually docked into the active sites based on site-directed mutagenesis information,^[30,31,35,61,62] as well as on the vicinity of the sites of metabolism towards the oxygen–iron system of the heme.

The docking algorithm used random orientations of the global minimum conformer as a starting point. To facilitate the automatic search for reasonable binding poses, besides these random orientations a pharmacophore-based alignment tool (Alignator)^[28] was employed. Alignator takes advantage of the conformer pool generated in the abovementioned conformational search. Each accepted conformer was aligned to a template molecule (for CYP2D6 a single and for CYP2C9 four templates were used) based on the matching pharmacophores. Of all possible solutions, only those having a maximum number of superimposable pharmacophores were retained. Alignment solutions that lead to ligand bumps with unmovable parts of the protein were filtered out (in this concept, only a locally induced fit related to the side-chain movement is supported. Heme, backbone and C β atoms, as well as bridged cysteine and proline residues are considered rigid).

In the next step, the residues in the active site (that have flexible side-chain atoms within 6.0 \AA from the ligand) were subjected to conformational analyses. For each side chain, only sterically allowed low-energy conformations were saved. A comprehensive combinatorial algorithm was employed to identify the best arrangement of all side chains that have a minimal number of bumps and the best hydrogen-bond interaction energy with the ligand and with each other. Up to four protein–ligand complexes with protein side chains that were reorganized for an optimal fit were prepared based on the following criteria (Figure 5, Complex A–D): A) the best root-mean-square deviation (rmsd) with the template, B) the best rmsd with the template, without bumps into unmovable protein atoms, C) the best rmsd with the template without bumps into unmovable backbone and side-chain atoms, and D) the best volume overlap with the template. These selection criteria ensure that both similar as well as rather different com-

pounds are rationally prealigned in the active site facilitating the subsequent automatic docking procedure.

To simulate local induced fit in Cheetah (formerly Yeti-Autodock) fully,^[17,23,27] an extended active-site region (any atom of amino acid residue located within 12 Å of any atom of the ligand molecule) is considered flexible. All poses (ligand conformations, orientations, positions) within 25% of the lowest-energy pose (typically 8–15 kcal mol⁻¹) featuring a minimal rmsd > 1.0 Å were composed in a 4D data set and forwarded to mQSAR for quantification of the binding affinity. The flow chart of the prealignment/docking along with the applied Monte Carlo protocol are shown in Figure 5.

Multidimensional QSAR: The employed mixed-model approach combining automated, flexible docking and multidimensional QSAR yielded more reliable binding-energy predictions than a direct scoring of the ligand–protein interactions of the experimental structure.^[63]

Quasar is a quasi-atomistic modeling concept that allows for mQSAR. The details of the software are published elsewhere,^[12,13] and are therefore only briefly summarized here. In Quasar, the binding site of the protein is represented by a surrogate, which consists of a three-dimensional surface at the van der Waals distance, surrounding the ligands that are superimposed in their bioactive conformation (as obtained, for example, from docking studies at the true biological receptor). The topology of this surface mimics the shape of the active site. This surface is then populated with quasi-atomistic properties, taking into account the potential of an amino acid to form a salt bridge and to act as hydrogen-bond donor and/or acceptor. Hydrogen-bond flip-flop, solvent, and contributions from neutral, positively charged, negatively charged, and hydrophobic residues are also accounted for.^[13] In addition to accepting 4D compound input (conformations, poses, protonation states, tautomers), Quasar allows the simulation of induced fit (corresponding to side-chain flexibility and moderate backbone motion at the true biological receptor), whereby six different protocols are evaluated simultaneously (5D-QSAR).^[13] The model family, typically consisting of 200 models, is evolved by using a genetic algorithm and provides an averaged prediction for each compound along with the variation over the model family. Quasar employs the following scoring function derived from the directional Yeti force field [Eq. (2)] in which $E_{\text{ligand-receptor}}$ is defined by [Eq. (3)].^[59]

$$E_{\text{binding}} = E_{\text{ligand-receptor}} - E_{\text{ligand desolvation}} - E_{\text{ligand internal strain}} - T\Delta S - E_{\text{induced fit}}$$

$$E_{\text{ligand-receptor}} = E_{\text{electrostatic}} + E_{\text{van der Waals}} + E_{\text{hydrogen bonding}} + E_{\text{polarization}}$$

In our simulations, ligands were represented as 4D ensembles. As a selection criteria for the training/test set splitting, the chemical diversity of the compounds was applied yielding a 46:10 (training/test set) ratio for CYP2D6 and 68:17 for CYP2C9, respectively. The compounds were grouped aiming for maximal diversity in terms of chemical scaffold and assigning the most and least active compounds to the training set. A family of 200 active-site models for each of the two cytochromes was evolved for 180 generations (36 000 crossovers) in the case of CYP2D6 and 40 generations (8000 crossovers) for CYP2C9.

Acknowledgement

We wish to express our gratitude to Dr. Daniel R. McMasters for kindly providing us with the affinities for the compounds used in

the QSAR study. Financial support by the Swiss National Foundation is gratefully acknowledged (grant number: 200021_119817).

Keywords: computational chemistry · cytochromes · enzymes · mQSAR · pharmacophore alignment

- [1] H. K. Kroemer, M. Eichelbaum, *Life Sci.* **1995**, *56*, 2285–2298.
- [2] T. H. Sullivan-Klose, B. I. Ghanayem, D. A. Bell, Z. Y. Zhang, L. S. Kaminsky, G. M. Shenfield, J. O. Miners, D. J. Birkett, J. A. Goldstein, *Pharmacogenetics* **1996**, *6*, 341–349.
- [3] C. A. Kemp, J. D. Marechal, M. J. Sutcliffe, *Arch. Biochem. Biophys.* **2005**, *433*, 361–368.
- [4] P. Rowland, F. E. Blaney, M. G. Smyth, J. J. Jones, V. R. Leydon, A. K. Oxbrow, C. J. Lewis, M. G. Tennant, S. Modi, D. S. Eggleston, R. J. Chenery, A. M. Bridges, *J. Biol. Chem.* **2006**, *281*, 7614–7622.
- [5] M. R. Wester, J. K. Yano, G. A. Schoch, C. Yang, K. J. Griffin, C. D. Stout, E. F. Johnson, *J. Biol. Chem.* **2004**, *279*, 35630–35637.
- [6] P. A. Williams, J. Cosme, A. Ward, H. C. Angova, D. M. Vinkovic, H. Jhoti, *Nature* **2003**, *424*, 464–468.
- [7] P. R. Porubsky, K. M. Meneely, E. E. Scott, *J. Biol. Chem.* **2008**, *283*, 33698–33707.
- [8] A. Seifert, S. Tatzel, R. D. Schmid, J. Pleiss, *Proteins Struct. Funct. Bioinf.* **2006**, *64*, 147–155.
- [9] M. de Groot, M. Ackland, V. Horne, A. Alex, B. Jones, *J. Med. Chem.* **1999**, *42*, 1515–1524.
- [10] S. Ekins, M. J. de Groot, J. P. Jones, *Drug Metab. Dispos.* **2001**, *29*, 936–944.
- [11] D. R. McMasters, R. A. Torres, S. J. Crathern, D. L. Dooney, R. B. Nachbar, R. P. Sheridan, K. R. Korzekwa, *J. Med. Chem.* **2007**, *50*, 3205–3213.
- [12] A. Vedani, M. Dobler, *J. Med. Chem.* **2002**, *45*, 2139–2149.
- [13] A. Vedani, M. Dobler, *Quant. Struct.-Acti. Relat.* **2002**, *21*, 382–390.
- [14] A. Vedani, M. Smiesko, M. Spreafico, O. Peristera, M. Dobler, *ALTEX* **2009**, *26*, 167–176.
- [15] M. A. Lill, F. Winiger, A. Vedani, B. Ernst, *J. Med. Chem.* **2005**, *48*, 5666–5674.
- [16] A. Vedani, M. Dobler, M. Lill, *Basic Clin. Pharmacol. Toxicol.* **2006**, *99*, 195–208.
- [17] A. Vedani, M. Dobler, M. Lill, *J. Med. Chem.* **2005**, *48*, 3700–3703.
- [18] M. Spreafico, B. Ernst, M. Lill, M. Smiesko, A. Vedani, *ChemMedChem* **2009**, *4*, 100–109.
- [19] M. Spreafico, M. Smiesko, O. Peristera, G. Rossato, A. Vedani, *Mol. Inf.* **2010**, *29*, 27–36.
- [20] O. Peristera, M. Spreafico, M. Smiesko, B. Ernst, A. Vedani, *Toxicol. Lett.* **2009**, *189*, 219–224.
 - [21] A. Vedani, A. Descloux, M. Spreafico, B. Ernst, *Toxicol. Lett.* **2007**, *173*, 17–23.
 - [22] A. Vedani, M. Zumstein, M. Lill, B. Ernst, *ChemMedChem* **2007**, *2*, 78–87.
 - [23] M. A. Lill, M. Dobler, A. Vedani, *ChemMedChem* **2006**, *1*, 73–81.
 - [24] J. Hritz, A. de Ruiter, C. Oostenbrink, *J. Med. Chem.* **2008**, *51*, 7469–7477.
- [25] F. P. Guengerich, *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 13565–13566.
- [26] M. J. de Groot, M. J. Ackland, V. A. Horne, A. A. Alex, B. C. Jones, *J. Med. Chem.* **1999**, *42*, 4062–4070.
- [27] <http://www.biograf.ch/downloads/VirtualToxLab.pdf>.
- [28] M. Smiesko, A. Vedani; *unpublished results*.
- [29] D. Zhang, J. P. Freeman, J. B. Sutherland, A. E. Walker, Y. Yang, C. E. Cer-niglia, *Appl. Environ. Microbiol.* **1996**, *62*, 798–803.
- [30] M. Paine, L. McLaughlin, J. Flanagan, C. Kemp, M. Sutcliffe, G. Roberts, C. Wolf, *J. Biol. Chem.* **2003**, *278*, 4021–4027.
- [31] J. Flanagan, J. Maréchal, R. Ward, C. Kemp, L. McLaughlin, M. Sutcliffe, G. Roberts, M. Paine, C. Wolf, *Biochem. J.* **2004**, *380*, 353–360.
- [32] I. Hanna, M. Kim, F. Guengerich, *Arch. Biochem. Biophys.* **2001**, *393*, 255–261.
- [33] M. Dachtler, H. Händel, T. Glaser, D. Lindquist, R. M. Hawk, C. N. Karson, R. A. Komoroski, K. Albert, *Magn. Reson. Chem.* **2000**, *38*, 951–956.
- [34] H. Yamazaki, T. Shimada, *Arch. Biochem. Biophys.* **1997**, *346*, 161–169.
- [35] K. Takashi, H. Tainaka, K. Kobayashi, T. Yasumori, M. Hosakawa, K. Chiba, *Pharmacogenetics* **2000**, *10*, 95–104.

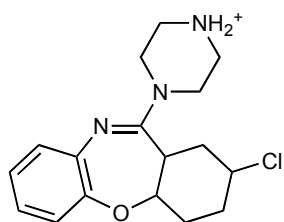
- [36] H. Shen, M. He, H. Liu, S. Wrighton, L. Wang, B. Guo, C. Li, *Drug Metab. Dispos.* **2007**, *35*, 1292–1300.
- [37] G. Mannens, C. Snel, J. Hendrickx, T. Verhaeghe, L. Le Jeune, W. Bode, L. van Beijsterveldt, K. Lavrijsen, J. Leempoels, N. Van Osselaer, A. Van Peer, W. Meuldermans, *Drug Metab. Dispos.* **2002**, *30*, 553–563.
- [38] K. Kramer Nielsen, K. Brösen, L. F. Gram, *Eur. J. Clin. Pharmacol.* **1992**, *43*, 405–411.
- [39] I. Zamora, L. Afzelius, G. Cruciani, *J. Med. Chem.* **2003**, *46*, 2313–2324.
- [40] S. Wold, L. Eriksson in *Chemometric Methods in Molecular Design (Methods and Principles in Medicinal Chemistry)*, (Ed.: H. van de Waterbeemd), Wiley-VCH, Weinheim, **1995**, pp. 309–318.
- [41] B. A. Hamelin, A. Bouayad, B. Drolet, A. Gravel, J. Turgeon, *Drug Metab. Dispos.* **1998**, *26*, 536–539.
- [42] G. Strobl, S. von Kruedener, J. Stöckigt, F. Guengerich, T. Wolff, *J. Med. Chem.* **1993**, *36*, 1136–1145.
- [43] S. Rao, R. Aoyama, M. Schrag, W. Trager, A. Rettie, J. Jones, *J. Med. Chem.* **2000**, *43*, 2789–2796.
- [44] M. Spreafico, M. Smiesko, G. Rossato, B. Ernst, A. Vedani; *unpublished results*.
- [45] L. Y. Lin, E. W. Di Stefano, D. A. Schmitz, L. Hsu, S. W. Ellis, M. S. Lennard, G. T. Tucker, A. K. Cho, *Drug Metab. Dispos.* **1997**, *25*, 1059–1064.
- [46] S. Preskorn, *J. Psychiatr. Pract.* **2003**, *9*, 150–158.
- [47] X. Li, T. Andersson, M. Ahlström, L. Weidolf, *Drug Metab. Dispos.* **2004**, *32*, 821–827.
- [48] T. Wolff, L. M. Distlerath, M. T. Worthington, J. D. Groopman, G. J. Hammons, F. F. Kadlubar, R. A. Prough, M. V. Martin, F. P. Guengerich, *Cancer Res.* **1985**, *45*, 2116–2122.
- [49] S. Islam, C. Wolf, M. Lennard, M. Sternberg, *Carcinogenesis* **1991**, *12*, 2211–2219.
- [50] L. Terfloth, B. Bienfait, J. Gasteiger, *J. Chem. Inf. Model.* **2007**, *47*, 1688–1701.
- [51] N. Manga, J. Duffy, P. Rowe, M. Cronin, *SAR QSAR Environ. Res.* **2005**, *16*, 43–61.
- [52] P. Zbinden, M. Dobler, G. Folkers, A. Vedani, *Quantitative Struct.-Act. Relat.* **1998**, *17*, 122–130.
- [53] K. J. Bowers, E. Chow, X. Huang, R. O. Dror, M. P. Eastwood, B. A. Grogens, J. L. Klepeis, I. Kolossvary, M. A. Moraes, F. D. Sacerdoti, J. K. Salmon, Y. Shan, D. E. Shaw, *Supercomputing (SC) 2006 Conference, Proceedings of the ACM/IEEE*, **2006**, 13 pp., CD-ROM.
- [54] A. D. McLachlan, *Acta Crystallogr. Sect. A* **1982**, *38*, 871–873.
- [55] W. L. Jorgensen, *J. Am. Chem. Soc.* **1988**, *110*, 1657–1666.
- [56] <http://www.schrodinger.com>.
- [57] C. J. Cramer, D. G. Truhlar, *J. Comput.-Aided Mol. Des.* **1992**, *6*, 629–666.
- [58] S. J. Weiner, P. A. Kollman, D. A. Case, U. C. Singh, C. Ghio, G. Alagona, S. Profeta, P. Weiner, *J. Am. Chem. Soc.* **1984**, *106*, 765–784.
- [59] A. Vedani, D. W. Huhta, *J. Am. Chem. Soc.* **1990**, *112*, 4759–4767.
- [60] V. Kumar, D. A. Rock, C. J. Warren, T. S. Tracy, J. L. Wahlstrom, *Drug Metab. Dispos.* **2006**, *34*, 1903–1908.
- [61] J. Flanagan, L. McLaughlin, M. Paine, M. Sutcliffe, G. Roberts, C. Wolf, *Biochem. J.* **2003**, *370*, 921–926.
- [62] A. Melet, N. Assrir, P. Jean, M. Pilar Lopez-Garcia, C. Marques-Soares, M. Jaouen, P. Dansette, M. Sari, D. Mansuy, *Arch. Biochem. Biophys.* **2003**, *409*, 80–91.
- [63] A. Vedani, M. Smiesko in *Computational Structural Biology: Methods and Applications*, (Eds.: T. Schwede, M. Peitsch), World Scientific Publishing Co. Pte. Ltd., London, **2008**, pp. 549–572.

Received: August 23, 2010
Published online on October 29, 2010

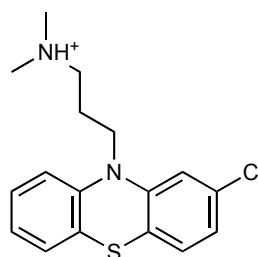
Appendix 2a: ChemMedChem 2010 – Supporting Information

Table 1: Structures of ligands binding to CYP2D6 and shared with CYP2C9 (*)

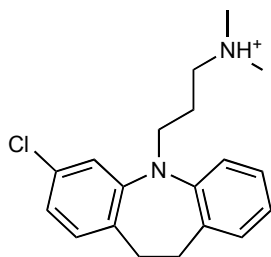
1 – Amoxapine (*)



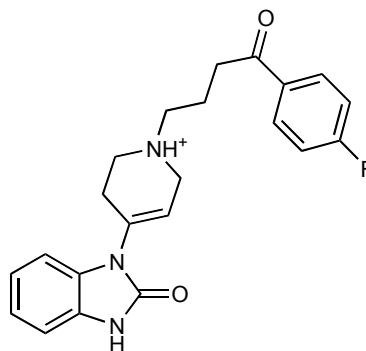
2 – Chlorpromazine (*)



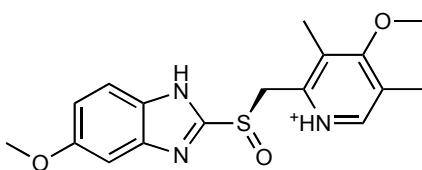
3 – Clomipramine (*)



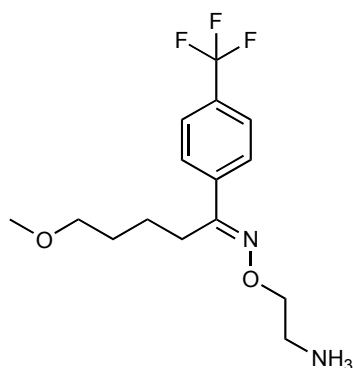
4 – Droperidol (*)



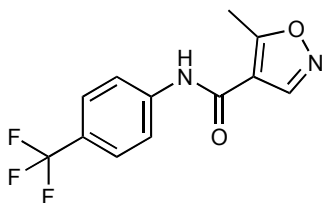
5 – Esomeprazole (*)



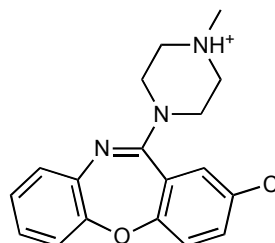
6 – Fluvoxamine (*)



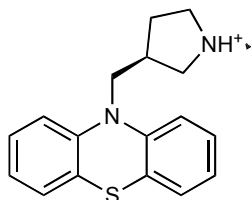
7 – Leflunomide (*)



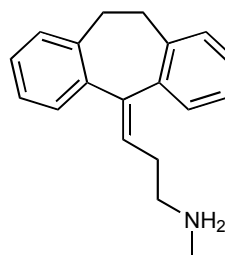
8 – Loxapine (*)



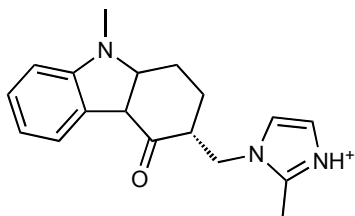
9 – Methdilazine (*)



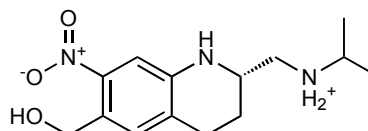
10 – Nortriptyline (*)



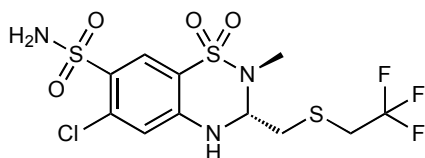
11 – Ondansetron (*)



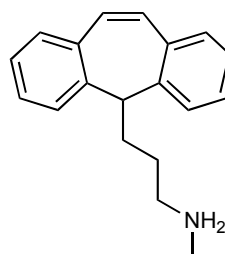
12 – Oxamniquine (*)



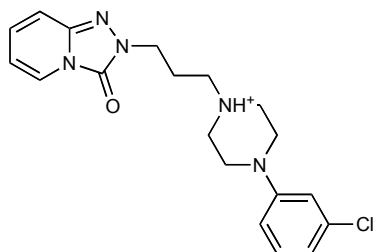
13 – Polythiazide (*)



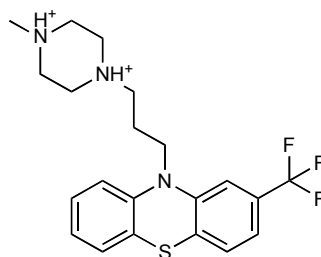
14 – Protriptyline (*)



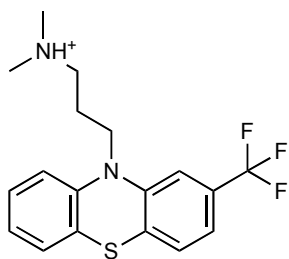
15 – Trazodone (*)



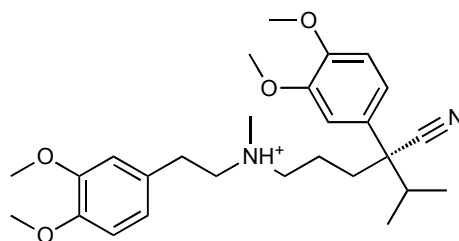
16 – Trifluoperazine (*)



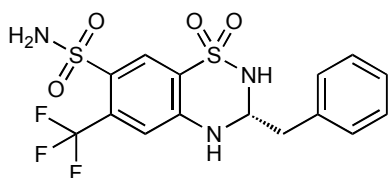
17 – Triflupromazine (*)



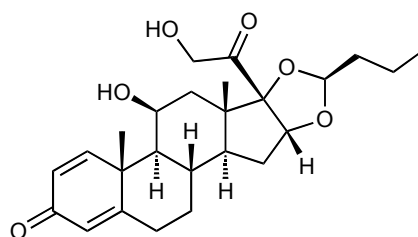
18 – Verapamil (*)



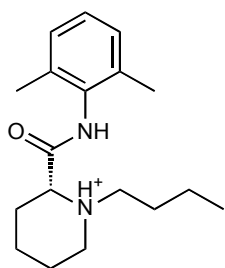
19 – Bendroflumethiazide



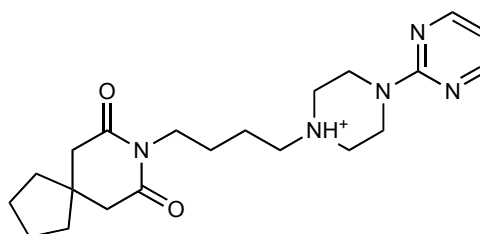
20 – Budesonide



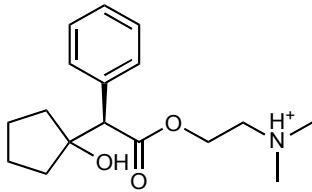
21 – Bupivacaine



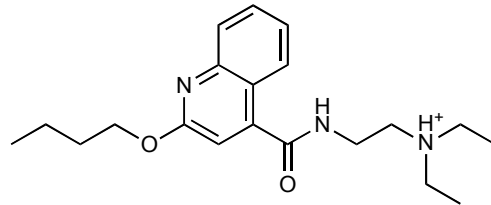
22 – Buspirone



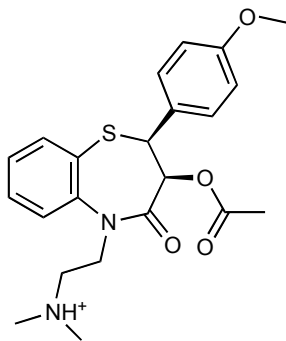
23 – Cyclopentolate



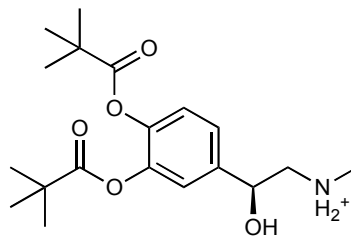
24 – Dibucaine



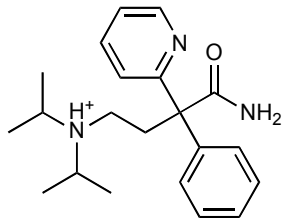
25 – Diltiazem



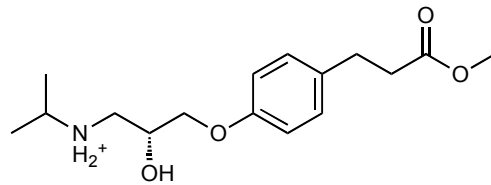
26 – Dipivefrim



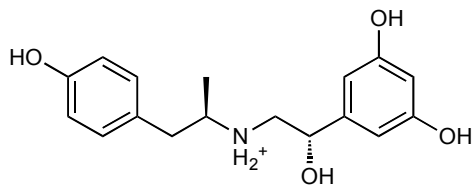
27 – Disopyramide



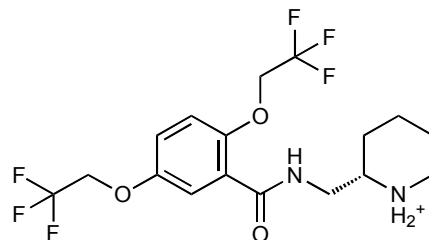
28 – Esmolol



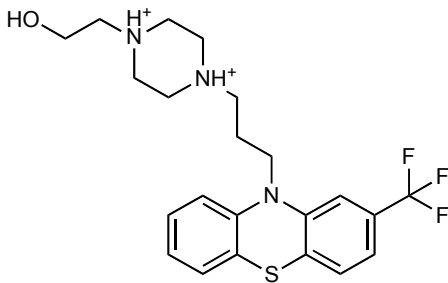
29 – Fenoterol



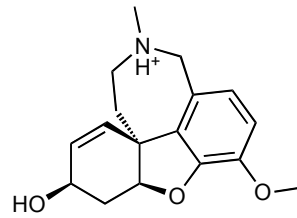
30 – Flecainide



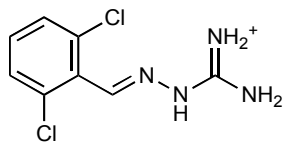
31 – Fluphenazine



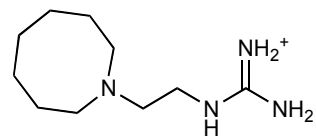
32 – Galantamine



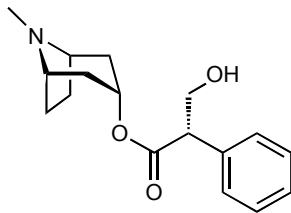
33 – Guanabenz



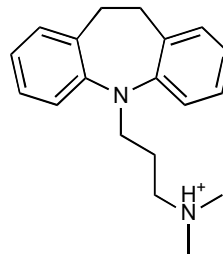
34 – Guanethidine



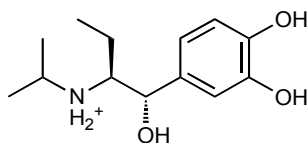
35 – Hyoscyamine



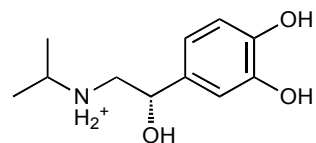
36 – Imipramine Hydrochloride



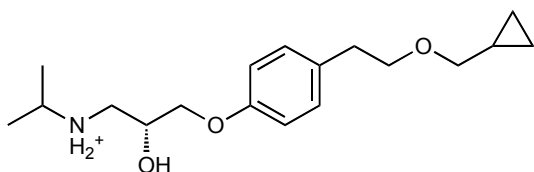
37 – Isoetharine



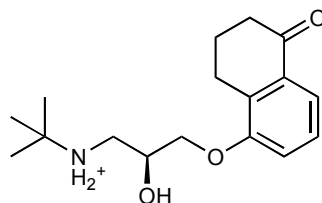
38 – Isoproterenoloxycodone



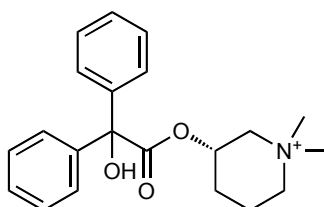
39 – Levobetaxol



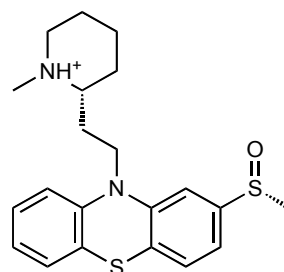
40 – Levobunol



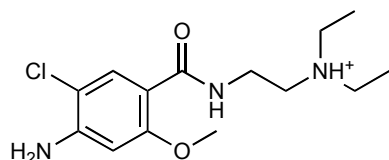
41 – Mepenzolate



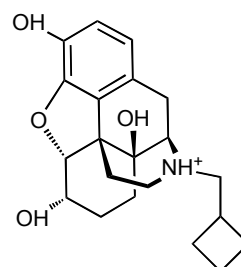
42 – Mesoridazine



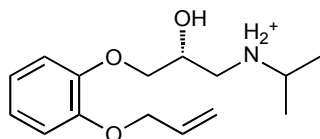
43 – Metoclopramide



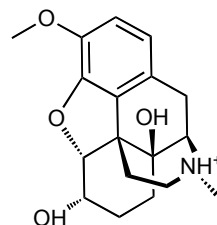
44 – Nalbuphine



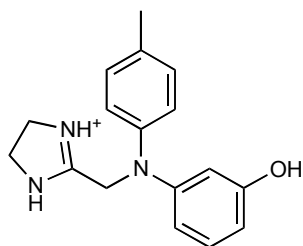
45 – Oxprenolol



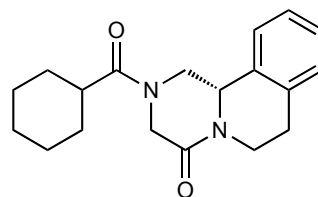
46 – Oxycodone



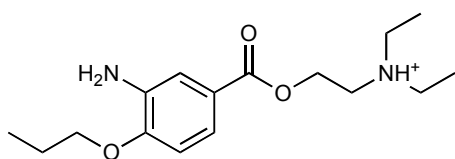
47 – Phentolamine



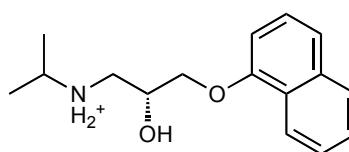
48 – Praziquantel



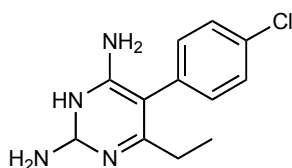
49 – Proparacaine



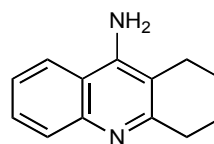
50 – Propranolol



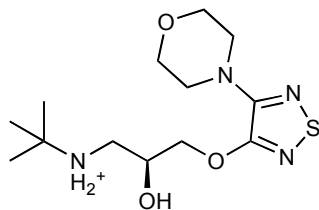
51 – Pyrimethamine



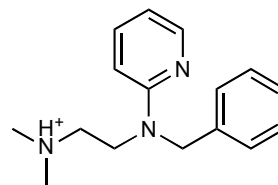
52 – Tacrine



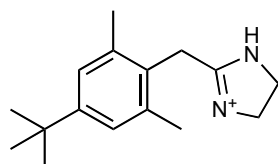
53 – Timolol



54 – Tripeleminamine



55 – Xylomethazoline



56 – Yohimbine

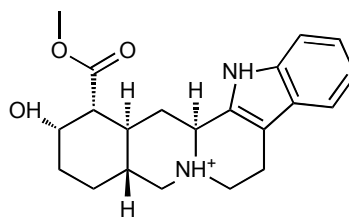
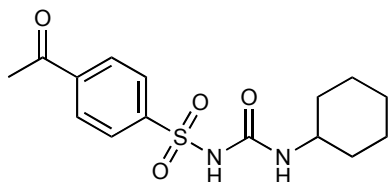
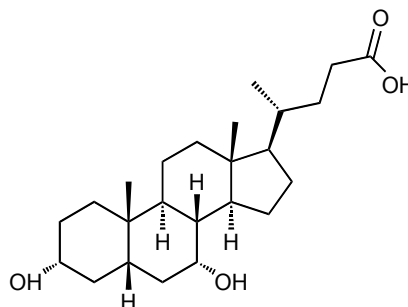


Table 2 : Structures of ligand binding to 2C9 and not shared with CYP2D6.

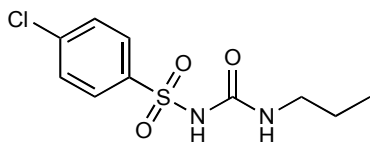
1 – Acetohexamide



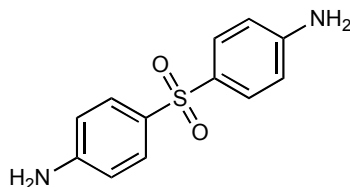
2 – Chenodiol



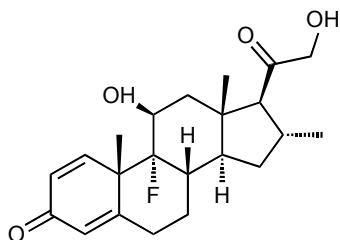
3 – Chlorpropamide



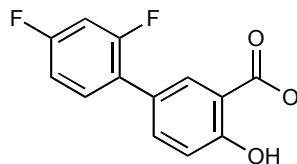
4 – Dapsone



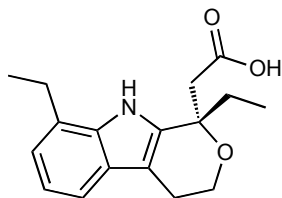
5 – Desoximetasone



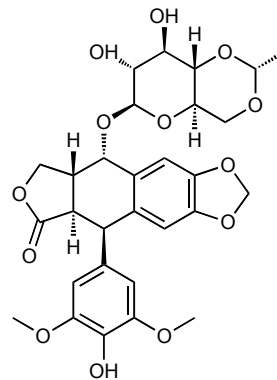
6 – Diflunisal



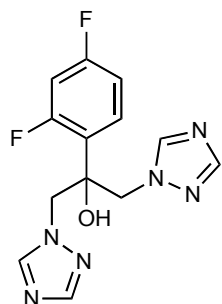
7 – Etodolac



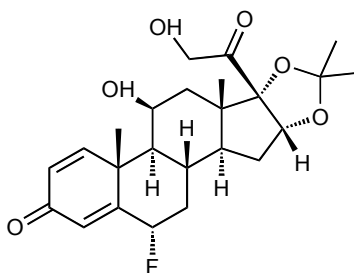
8 – Etoposide



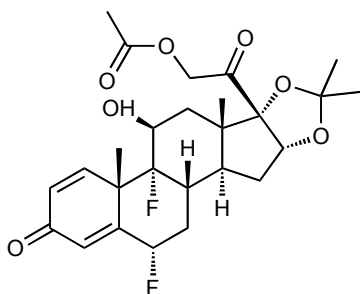
9 – Fluconazole



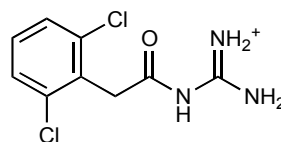
10 – Flunisolide



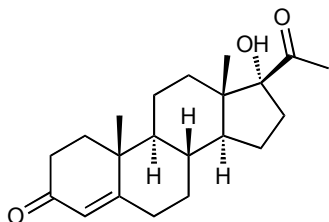
11 – Flucicnonide



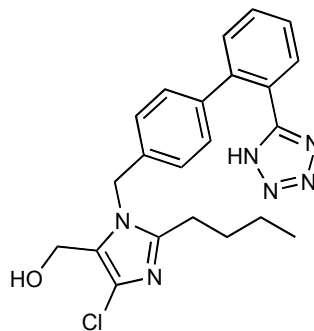
12 – Guanfacine



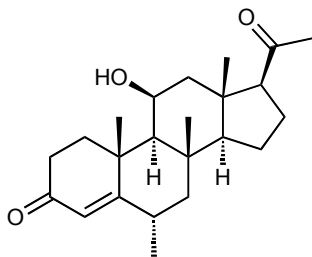
13 – Hydroxyprogesterone



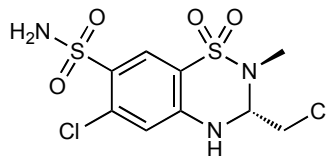
14 – Losartan



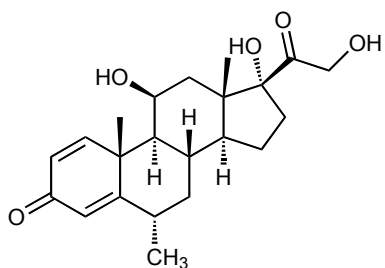
15 – Medrisone



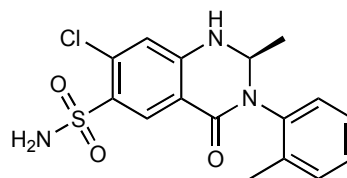
16 – Methyclothiazide



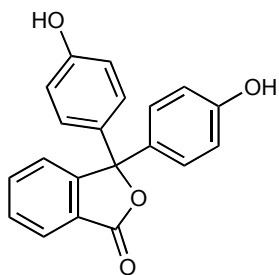
17 – Methylprednisolone



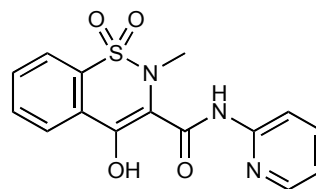
Metolazone



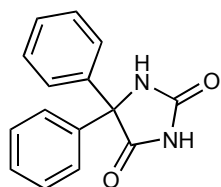
19 – Phenolphthalein



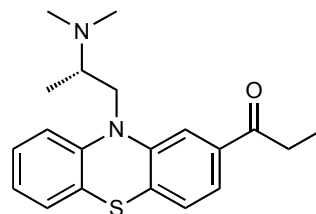
20 – Piroxicam



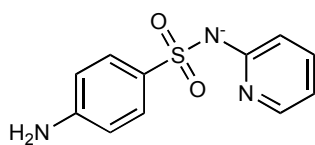
21 – Phenytoin



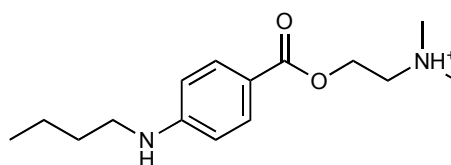
22 – Propiomazine



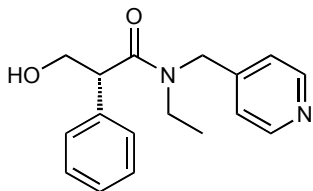
23 – Sulfapyridine



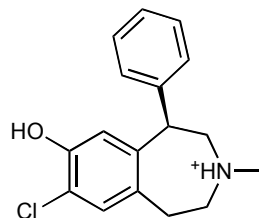
24 – Tetracaine



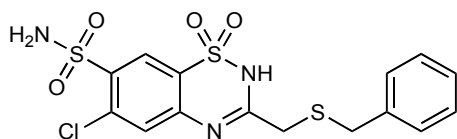
25 – Tropicamide



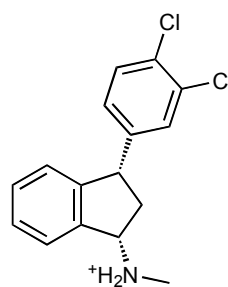
26 – MFCD00069249



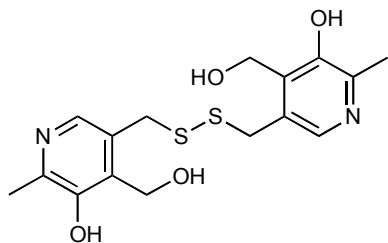
27 – MFCD00078969



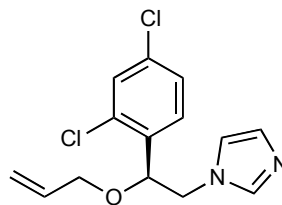
28 – MFCD00083178



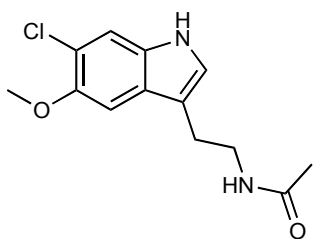
29 – MFCD00151477



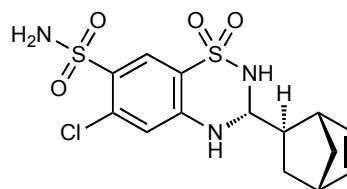
30 – MFCD00186378



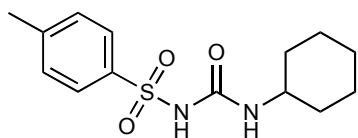
31 – MFCD00209923



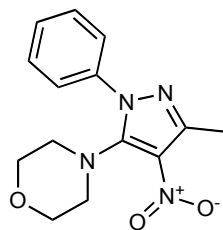
32 – MFCD00210192



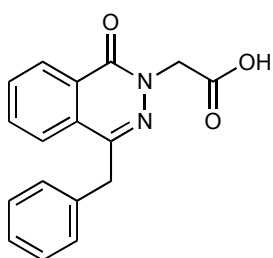
33 – MFCD00242887



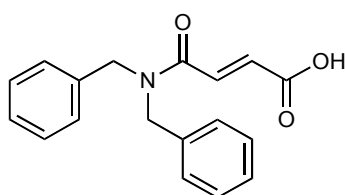
34 – MFCD00277611



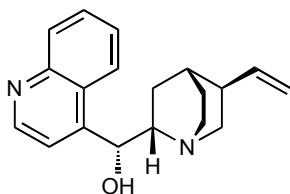
35 – MFCD00406072



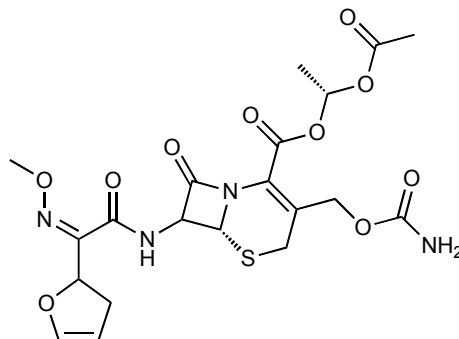
36 – MFCD00588221



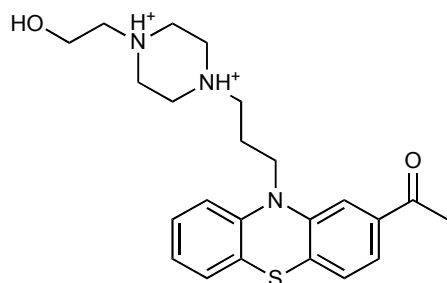
37 – MFCD00778684



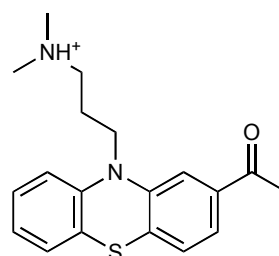
38 – MFCD00864991



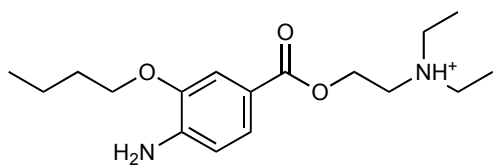
39 – MFCD00866682



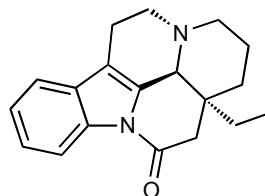
40 – MFCD00867988



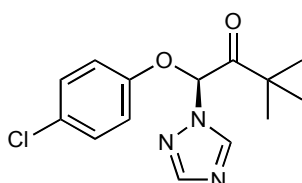
41 – MFCD01102908



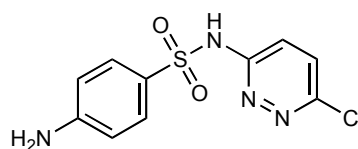
42 – MFCD01456444



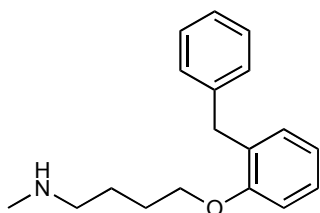
43 – MFCD02060107



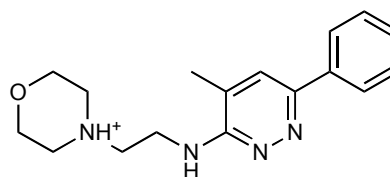
44 – MFCD02668080



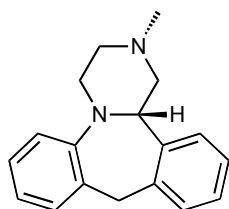
45 – MDDR_090818



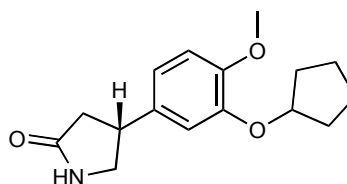
46 – MDDR_091137



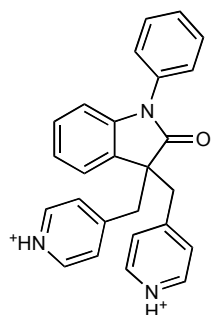
47 – MDDR_091446



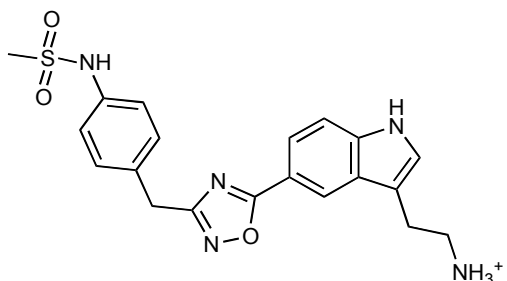
48 – MDDR_107859



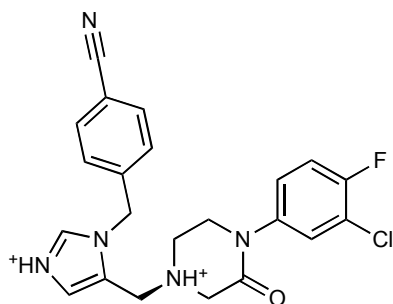
49 – MDDR_137340



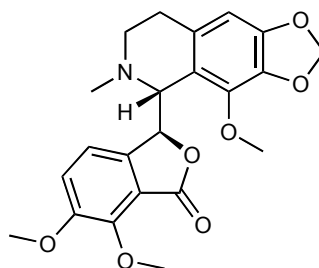
50 – MDDR_205042



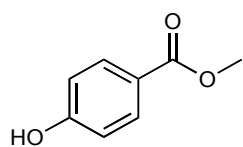
51 – MDDR_274259



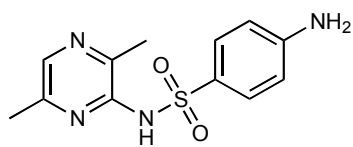
52 – MDDR_274460



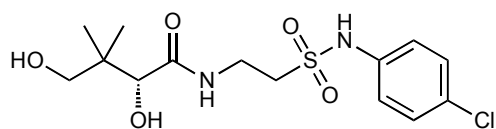
53 – NCI_3827



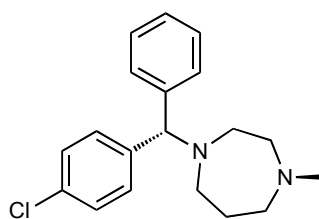
54 – NCI_13149



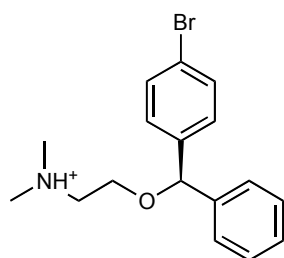
55 – NCI_23822



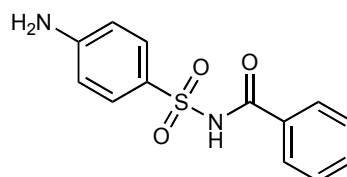
56 – NCI_25132



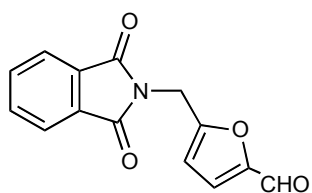
57 – NCI_36113



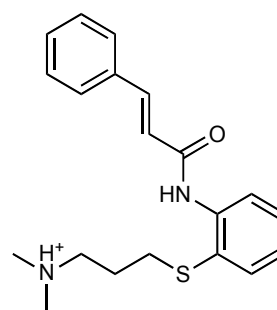
58 – NCI_74587



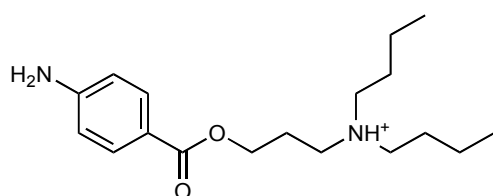
59 – NCI_119558



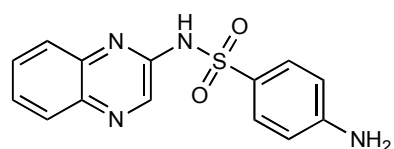
60 – NCI_125717



61 – NCI_403903



62 – NCI_683543



63 – NCI_683544

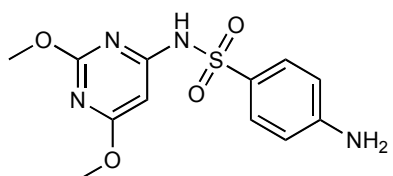


Table 3: List of twelve compounds binding to CYP2C9 and comparison of different binding assay values (Kumar et al.). In the column titled (Recombinant Enzymes) the values of K_i as from the dataset used to build the models (McMasters et al.).

Ligand Name	Recombinant Enzymes (μM)	Supersomes (μM)	HLM(μM)	RECO (μM)
Benzbromarone		0.004	0.011	0.11
Diclofenac		1.6	2.9	9.3
<u>Fluvoxamine</u>	<u>4.8</u>	<u>0.63</u>	<u>5.7</u>	<u>6.6</u>
Gemfibrozil		12	0.98	14
S-Ibuprofen		4.3	3.2	17
Indomethacin		53	13	49
<i>Ketoconazole</i>		<i>0.38</i>	<i>2.1</i>	<i>2.4</i>
<i>Miconazole</i>		<i>0.028</i>	<i>0.038</i>	<i>0.11</i>
<u>Phenytoin</u>	<u>68</u>	<u>5.6</u>	<u>10</u>	<u>11</u>
<u>Piroxicam</u>	<u>20</u>	<u>0.47</u>	<u>9.8</u>	<u>6.7</u>
Sulfaphenazole		0.06	0.15	1.3
Tolbutamide		1.5	3.5	2

Table 3a: Ratio of the K_i values obtained from the *recombinant enzymes* with supersomes, HLM and RECO (a purified, reconstituted preparation) for the three reference compounds used to compared the affinity values. In bold the binding assay displaying minor maximal deviance from the values of the McMasters dataset.

	Ratio Recombinant Enzymes / Supersomes	Ratio Recombinant Enzymes / HLM	Ratio Recombinant Enzymes/ RECO
<u>Fluvoxamine</u>	<u>7.60</u>	<u>0.84</u>	<u>0.73</u>
<u>Phenytoin</u>	<u>12.1</u>	<u>6.63</u>	<u>6.09</u>
<u>Piroxicam</u>	<u>42.4</u>	<u>2.04</u>	<u>2.98</u>

Table 4: List of software and methods used to generate the models. Further information for each method are provided.

<p>Protein Relaxation – MD – (Desmond)</p>	<ul style="list-style-type: none"> • PDB codes: 2F9Q (2D6) 1R9O (2C9) • Solvent Model : TIP3P • Relaxation Protocol : 72 ps • Simulation Length : 1 ns • Salt Concentration : 0.15 M
<p>Conformers Generation (MacroModel)</p>	<ul style="list-style-type: none"> • Protonation/Tautomeric State assignment (Epik) • Conformational Search : 5000 steps • Partial Charges Calculation: CM1 (AMSOL) • Solvation Energy Calculation: SM2 (AMSOL)
<p>Pharmacophore Pre-alignment (Alignator)</p>	<ul style="list-style-type: none"> • Pharmacophore Alignment • Active Site Rearrangement <p>Output A – Best template <i>rmsd</i> Output B – Best template <i>rmsd</i> with no main-chains bumps Output C – Best template <i>rmsd</i> with no side-chains bumps Output D – Best volume overlap</p>
<p>Automated Flexible Docking (Cheetah)</p>	<ul style="list-style-type: none"> • Configuration Generation <ul style="list-style-type: none"> - Torsional flipping (local/global, uptree/ dountree) - Local rigid-body rotation & translation - <i>Monte-Carlo</i> Metropolis algorithm • Energy Refinement of Active Site area: 12 Å • Global Poses Generation
<p>Ranking/Scoring of Poses</p>	<ul style="list-style-type: none"> • Ligand conformations, orientations, positions within 25% from the lowest-energy pose (~ 8-15 kcal/mol) • Minimal <i>rmsd</i> difference : 1 Å
<p>mQSAR (Quasar)</p>	<ul style="list-style-type: none"> • Models number: 200 • Receptor Surrogates (vdW distance) • Quasi-atomistic properties: (salt bridges, hydrogen bond donors/acceptors, hydrogen bond flip-flop, solvent and charged residues contribution) • mQSAR: different protocols evaluated

Table 5: Details of the predictive r^2 obtained from the Y-scrambling for CYP2D6.

#	Predictive r^2	#	Predictive r^2
1	-0.73	21	-0.42
2	-0.02	22	-0.19
3	-0.017	23	-0.27
4	-0.27	24	-0.87
5	-0.17	25	-0.23
6	-0.9	26	-0.042
7	-0.83	27	-0.36
8	-0.048	28	-0.45
9	-0.21	29	-0.81
10	0.24	30	-0.47
11	-1.46	31	-0.79
12	0.083	32	-0.54
13	-0.23	33	0.31
14	-0.077	34	-0.76
15	-0.17	35	0.075
16	-0.004	36	-0.47
17	-0.005	37	0.079
18	0.56	38	-0.61
19	-0.14	39	-1.21
20	-0.14	40	-0.14

Table 6: Details of the predictive r^2 obtained from the Y-scrambling for CYP2C9.

#	Predictive r^2	#	Predictive r^2
1	-1.0	21	0.42
2	-0.45	22	0.12
3	-0.14	23	0.57
4	-0.60	24	-0.48
5	-1.85	25	-0.50
6	-0.79	26	-0.007
7	-0.71	27	-0.009
8	-0.29	28	0.23
9	-0.26	29	-0.091
10	0.023	30	-0.31
11	-0.065	31	-0.47
12	-0.68	32	0.13
13	-0.15	33	-0.007
14	-0.78	34	-1.26
15	-0.21	35	-0.24
16	-0.23	36	0.41
17	-0.065	37	-0.12
18	-1.00	38	-0.84
19	-0.70	39	0.29
20	0.42	40	0.40

Appendix 3: Bio. & Med. Chem. 2010



Contents lists available at ScienceDirect

Bioorganic & Medicinal Chemistry

journal homepage: www.elsevier.com/locate/bmc

Design, synthesis, biological evaluation, and modeling of a non-carbohydrate antagonist of the myelin-associated glycoprotein

Oliver Schwardt^a, Hendrik Koliwer-Brandl^b, Raphael Zimmerli^a, Stefanie Mesch^a, Gianluca Rossato^a, Morena Spreafico^a, Angelo Vedani^a, Sørge Kelm^b, Beat Ernst^{a,*}

^aInstitute of Molecular Pharmacy, Pharmacenter, University of Basel, Klingelbergstrasse 50, CH-4056 Basel, Switzerland

^bDepartment of Physiological Biochemistry, University of Bremen, D-28334 Bremen, Germany

ARTICLE INFO

Article history:

Received 26 May 2010

Accepted 12 August 2010

Available online 17 August 2010

Keywords:

Carbohydrate mimetics

Siglecs

Myelin-associated glycoprotein (MAG)

Binding affinity

Hapten inhibition assay

Surface plasmon resonance (SPR)

Molecular modeling

Key polar groups

ABSTRACT

Broad modifications of various positions of the minimal natural epitope recognized by the myelin-associated glycoprotein (MAG), a blocker of regeneration of neurite injuries, produced sialosides with nanomolar affinities. However, important pharmacokinetic issues, for example, the metabolic stability of these sialosides, remain to be addressed. For this reason, the novel non-carbohydrate mimic **3** was designed and synthesized from (–)-quinic acid. For the design of **3**, previously identified beneficial modifications of side chains of Neu5Ac were combined with the replacement of the ring oxygen by a methylene group and the substitution of the C(4)-OH by an acetamide. Although docking experiments to a homology model of MAG revealed that mimic **3** forms all but one of the essential hydrogen bonds identified for the earlier reported lead **2**, its affinity was substantially reduced. Extensive molecular-dynamics simulation disclosed that the missing hydrogen bond of the former C(8)-OH leads to a change of the orientation of the side chain. As a consequence, an important hydrophobic contact is compromised leading to a loss of affinity.

© 2010 Elsevier Ltd. All rights reserved.

1. Introduction

Siglecs^{1,2} (sialic acid-binding immunoglobulin like-lectins) form a sub-group of the I-type lectins and function as cell signaling co-receptors primarily expressed on leukocytes to mediate acquired and innate immune functions. They can be divided into two subsets: the first, evolutionary conserved group consists of Siglec 1, 2, and 4, which show selective binding properties: Siglec 1 (also known as sialoadhesin) and Siglec 4 (myelin-associated glycoprotein, MAG) preferentially bind $\alpha(2 \rightarrow 3)$ -linked *N*-acetylneuraminic acid (Neu5Ac), whereas Siglec 2 (CD22) is highly specific for $\alpha(2 \rightarrow 6)$ -linked Neu5Ac. In contrast, members of the second Siglec 3-related group (Siglec 3 and Siglecs 5–13) are more promiscuous in their binding, often recognizing more than one presentation of Neu5Ac.

Abbreviations: DBU, 8-diazabicyclo[5.4.0]undec-7-en; DCM, dichloromethane; DMAP, 4-dimethylamino-pyridine; DMF, *N,N*-dimethylformamide; DMP, 2,2-dimethoxypropane; FAC, 2-fluoroacetyl; IgG, immunoglobulin G; K_D , dissociation constant; MAG, myelin-associated glycoprotein; Neu5Ac, *N*-acetylneuraminic acid; NgR, Nogo receptor; NMR, nuclear magnetic resonance; PDB, protein data bank; RP, reversed phase; SAR, structure-affinity relationship; SPR, surface plasmon resonance; TFA, trifluoroacetic acid; THF, tetrahydrofuran; *p*-Ts, *p*-tolylsulfonyl.

* Corresponding author. Tel.: +41 267 15 51; fax: +41 267 15 52.

E-mail address: beat.ernst@unibas.ch (B. Ernst).

0968-0896/\$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmc.2010.08.027

The most comprehensively characterized Siglecs are CD22, a regulatory protein on B lymphocytes that prevents the over-activation of the immune system and the development of autoimmune diseases, and MAG, one of several inhibitor proteins that block regeneration of injuries of the central nervous system (CNS).^{2–10} MAG is located on the surface of neurons and interacts with two classes of targets: proteins of the family of Nogo receptors (NgR)^{11,12} and brain gangliosides, specifically GD1a, GT1b, or GQ1b.^{2,13–15} Although the relative roles of gangliosides and NgRs as MAG ligands have yet to be resolved,^{10,16} in some systems, MAG inhibition is completely reversed by sialidase treatment, suggesting that MAG uses sialylated glycans as its major axonal ligands.¹⁷ Therefore, blocking MAG with potent antagonists may be a valuable therapeutic approach to enhance axon regeneration.¹⁸

SAR studies have revealed that the terminal tetrasaccharide epitope of GQ1b α (Fig. 1) shows superior binding to MAG compared with the terminal trisaccharide epitope present in GD1a or GT1b.^{19,20} Furthermore, the MAG-affinity of tetrasaccharide **1** (Fig. 1), could clearly be correlated with its ability to reverse MAG-mediated inhibition of axonal outgrowth.¹⁸

To reduce the structural complexity of tetrasaccharide **1** numerous MAG antagonists have been synthesized.^{21–23} Based on structural information obtained by trNOE NMR²⁴ and STD NMR,²⁵ the Gal $\beta(1 \rightarrow 3)$ GalNAc core was successfully replaced by non-carbohydrate linkers²² and the $\alpha(2 \rightarrow 6)$ -linked sialic acid by polar as

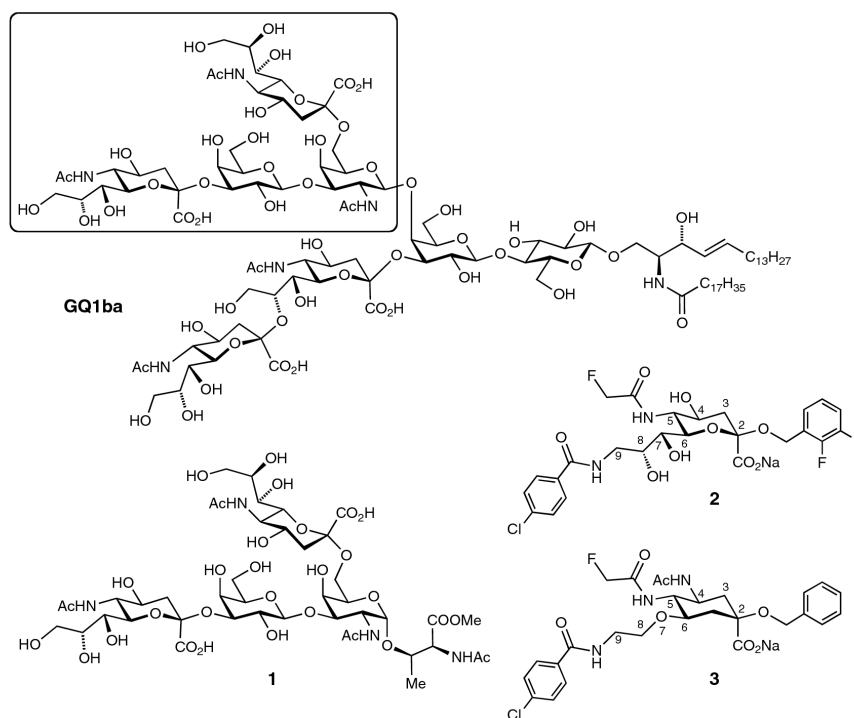


Figure 1. MAG antagonists; GQ1ba with the essential tetrasaccharide binding epitope highlighted, tetrasaccharide derivative **1**,¹⁸ sialic acid derivative **2**,³¹ and the target molecule of this communication, the non-carbohydrate mimic **3** (for better comparability, the carbon atoms in **3** are numbered in analogy to compound **2**).

well as lipophilic substituents.²³ A further simplification of lead structure **1** was reported by Kelm and Brossmer who found that sialic acid derivatives modified in the 2-, 5-, or 9-position exhibit enhanced antagonistic activity.^{26–28} Finally, when the best modifications found for the 2-,²⁹ 4-,³⁰ 5-,³¹ and 9-position³² were combined in one molecule, we identified the nanomolar low-molecular weight antagonist **2**.³¹

However, an important pharmacokinetic issue, namely the metabolic stability of sialosides like **2**, remains to be addressed. In general, the substrate specificity of mammalian sialidases is determined by the linkage type of the terminal sialic acid residue (2 → 3, 2 → 6, or 2 → 8) and does not depend on the structure of the underlying oligosaccharide.³³ Hence, a fast metabolic cleavage of sialosides of type **2** by sialidases cannot be excluded.

To solve this metabolic challenge, we replaced the pyranose core in **2** by a cyclohexane moiety. Here, we describe design, synthesis, and biological evaluation of the corresponding non-carbohydrate lead compound **3**. It still contains the beneficial modifications identified for the 2-, 5-, and 9-position (see compound **2**), however a simplified glycerol side chain.

2. Results and discussion

By manual docking of sialoside **2** (K_D 500 nM)³¹ to a homology model of MAG²³ the pharmacophores involved in this carbohydrate-lectin interaction were identified (see Fig. 3B). The most important contributions stem from a salt bridge formed by the carboxylic acid and Arg118^{26,34,35} and a hydrogen bond of the C(9)-NH with the backbone carbonyl of Thr128. Additionally, hydrogen-

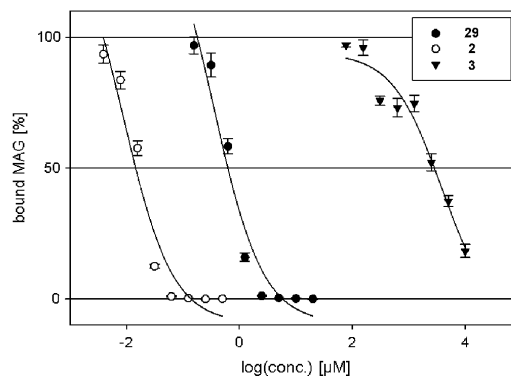


Figure 2. Hapten inhibition assay with sialosides **29** and **2** and non-carbohydrate mimic **3**. Fetuin-coated microtiter plates^{26,32,41} were incubated with antibody-complexed MAGd1–3-Fc in the presence of the indicated compounds at different concentrations. Binding and inhibition were determined as described for the hapten inhibition assay in Section 4.2.

bond formations of C(5)-NH with the backbone carbonyl of Gln126 and the backbone NH of Thr128 with the C(8)-OH were identified, whereas the C(4)- and C(7)-OH exhibit only minor contributions to binding. Overall, this hydrogen bond network nicely correlates with the SAR study published by Kelm and Schauer.²⁸ Furthermore, considerable contributions to the binding affinity result from two hydrophobic interactions: (i) the *p*-chlorobenzamide

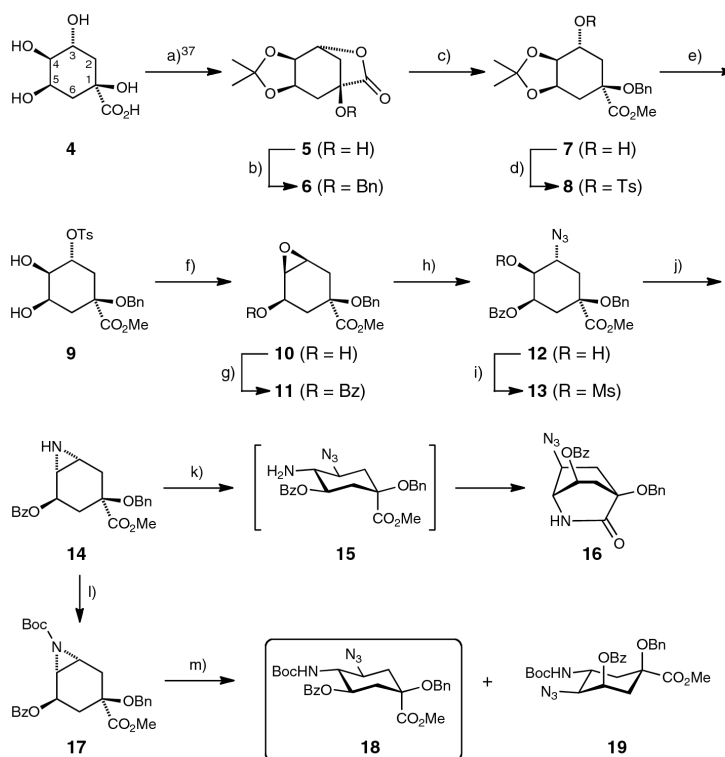
points into a hydrophobic pocket formed by Ser130 and the side chain methylenes of Glu131 and (ii) the 2,3-difluorobenzyl aglycone is hosted by a second hydrophobic pocket established by Trp59, Tyr60, Tyr69, and Tyr116. Most importantly, the pyranose ring oxygen was not found to play a significant role in the interaction of **2** and MAG.

Based on these findings, the non-carbohydrate mimic **3** (Fig. 1) was designed. Since the pyranose ring oxygen in **2** does not contribute to binding, it was exchanged by a methylene group. In addition, the hydroxyl group in the 4-position was replaced by an amide. By this modification, the H-bond with Gln126, which was shown to exhibit a minor contribution to binding,²⁸ is well preserved (see Fig. 3B and D). Furthermore, the additional acyl group allows the exploration of supplementary beneficial interactions with the target protein. Finally, the *p*-chlorobenzamido-substituted glycerol side chain was substituted by a simplified version, the 2-(*p*-chlorobenzamido)-ethoxy moiety. This modification maintains the essential H-bond originating from the C(9)-NH in **2**, but neglects the C(7)- and C(8)-OH, where at least the latter one was found to contribute to the binding to MAG.²⁸ However, the role of this functional group was only investigated with oligosaccharides and not in combination with other modifications in Neu5Ac derivatives,^{26–28} and therefore is difficult to assess quantitatively.

2.1. Synthesis of sialic acid mimic **3**

Because of the similar substitution pattern, that is, three ring substituents in the 4-, 5-, and 6-position and the α -hydroxy carboxylate function, (–)-quinic acid (**4**) is a feasible precursor for the synthesis of **3**. The main synthetic task is the inversion of the *trans*-C(3) and C(4) hydroxyls into the corresponding *trans*-C(3) and C(4) amino groups present in **3**.

The synthesis of the core building block **18** (Scheme 1) was performed in analogy to the synthesis of oseltamivir.^{36,37} Treatment of (–)-quinic acid (**4**) with 2,2-dimethoxypropane (DMP) in acetone yielded lactone **5** in 87%. Benzoylation of the free hydroxy function (\rightarrow **6**) and subsequent methanolysis generated methyl ester **7**. Sulfonic ester **9** was obtained almost quantitatively by tosylation of **7** followed by cleavage of the acetonide function in **8**. Treatment of **9** with DBU in THF led to epoxide **10** in 96%. Nucleophilic ring opening of the benzoyl protected epoxide **11** with sodium azide in the presence of ammonium chloride generated azido alcohol **12** in 78% yield. The high regioselectivity of the epoxide opening can be attributed to the steric as well as the electron withdrawing influence of the adjacent benzoate. Conversion of azide **12** to aziridine **14** was achieved by a two-step procedure: first, mesylation of the free hydroxyl group in **12** (\rightarrow **13**) and second, reduction of the azide using modified Staudinger conditions.³⁸ Unexpectedly, the



Scheme 1. Reagents and conditions: (a) DMP, cat. *p*-TsOH, acetone, 70 °C, 87%; (b) BnBr, NaH, DMF, 0 °C to rt, 94%; (c) NaOMe, MeOH, rt, 3 h, 73%; (d) TsCl, cat. DMAP, py, rt, 98%; (e) 80% aq AcOH, 60 °C, 3 h, 98%; (f) DBU, THF, 0 °C to rt, 18 h, 96%; (g) BzCl, cat. DMAP, py, 97%; (h) NaN₃, NH₄Cl, MeOH/H₂O (8:1), 70 °C, 78%; (i) MsCl, NEt₃, DCM, 0 °C to rt, 4 h, 94%; (j) PPh₃, THF then NEt₃, H₂O, rt, 5 h, 67%; (k) NaN₃, NH₄Cl, DMF, 65 °C, 50%; (l) Boc₂O, cat. DMAP, NEt₃, DCM, 0 °C to rt, 5 h, 96%; (m) NaN₃, NH₄Cl, DMF, 70 °C, **18**: 64%, **19**: 27%.

aziridine ring opening in **14** with sodium azide resulted exclusively in the bicyclic lactam **16**, resulting from a nucleophilic attack of the amine group in intermediate **15** to the carbonyl of the ester moiety. To avoid the lactame formation, the ring opening was performed with Boc-protected aziridine **17**. Again, an excellent chemical yield, but only a modest regioselectivity of 2.4:1 (**18**, 64% and **19**, 27%) was achieved.

When **18** was debenzoylated using Zemplén conditions, **20** was obtained in almost quantitative yield (Scheme 2). Since the subsequent allylation under phase transfer conditions afforded a mixture of allyl ester, methyl ester, and free acid, saponification with NaOH and re-esterification with TMS-CHN₂ was necessary to obtain uniform **21**. Conversion of its allyl group into the hydroxyethyl side chain (**22**) was accomplished in 65% yield by oxidative cleavage using osmium tetroxide and sodium periodate in the presence of 2,6-lutidine³⁹ followed by reduction with sodium borohydride. The azido-group in **22** was then selectively reduced by catalytic hydrogenation in the presence of triethylamine⁴⁰ followed by O- and N-acetylation (**23**). O-deacetylation (**24**), tosylation (**25**) and substitution of the tosylate by azide afforded **26**. The azido group was then converted into a *p*-chlorobenzamide (**27**) using the previously applied strategy of catalytic hydrogenation in the presence of NEt₃ and subsequent acylation. The conversion of the carbamate in **27** into a fluoroacetamido group via a two-step procedure (Boc-deprotection with TFA, followed by acylation) gave **28** in 69%. After saponification of the ester, the sodium salt of the final compound **3** was obtained by ion exchange chromatography in 83%.

2.2. Biological evaluation

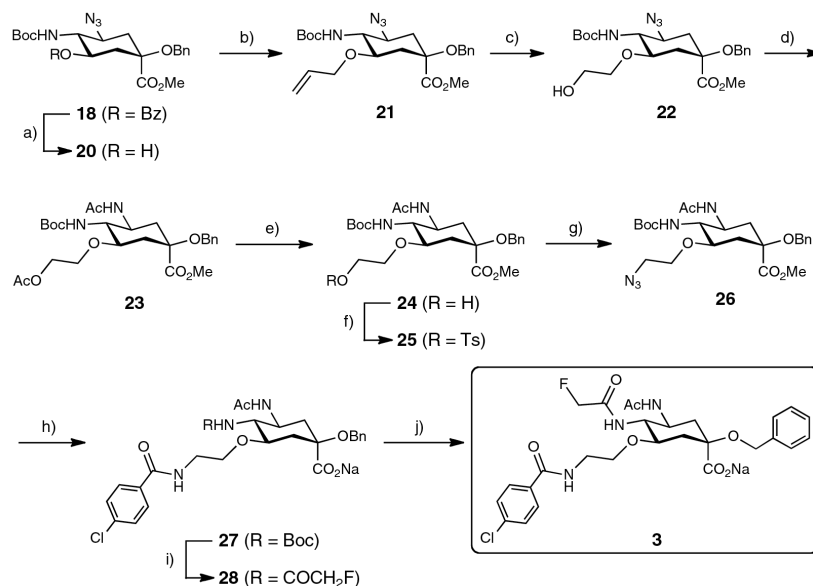
The interaction of MAG with sialoside **2** and the sialic acid mimetic **3** was analyzed by two previously reported assay formats; a fluorescent hapten binding assay⁴¹ and a surface plasmon reso-

nance (SPR) based biosensor (Biacore) experiment^{23,31} (Table 1). For the hapten inhibition assay, a recombinant protein consisting of the three N-terminal domains of MAG and the Fc part of human IgG (MAG_{d1-3}-Fc) was produced by expression in CHO-Lec1 cells and affinity purification on protein A-agarose.⁴¹ The relative inhibitory concentrations (rIC₅₀) of the test compounds were determined in microtiter plates coated with fetuin as binding target for MAG_{d1-3}-Fc. By complexing the Fc-part with alkaline phosphatase-labeled anti-Fc antibodies and measuring the initial velocity of fluorescein release from fluorescein diphosphate, the amount of bound MAG_{d1-3}-Fc could be determined. Four independent titrations were performed for each compound (Fig. 2). To ensure comparability with earlier results,^{31,32} the affinities were measured relative to the reference compound **29**³² (rIC₅₀ of 1, Table 1, entry 1). For the K_D determination in the Biacore assay, MAG_{d1-3}-Fc was immobilized on a dextran chip modified with covalently bound protein A. A reference cell providing only protein A was used to compensate for unspecific binding to the matrix.

Whereas sialoside **2** (rIC₅₀ 0.02, K_D 0.5 μM, entry 2) showed in both assays an approx. 50-fold improvement in affinity compared to the reference compound **29**, mimic **3** exhibited an IC₅₀ of approx. 3 mM (Fig. 2) in the hapten inhibition assay (rIC₅₀ 4300, entry 3), which corresponds to an about 200,000-fold reduced affinity for MAG compared to **2**. In the Biacore experiment compound **3** showed no binding up to a concentration of 800 μmol.

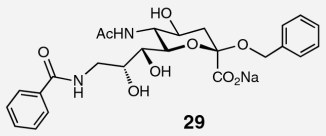
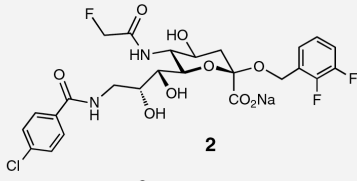
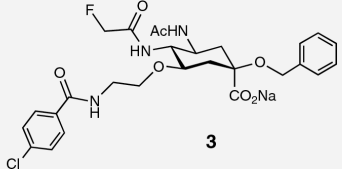
2.3. Molecular modeling

To elucidate this unexpected loss of binding affinity for mimic **3**, extensive molecular docking studies were performed. Since no crystal structure of MAG is yet available, the previously described homology model of the ligand-binding domain of MAG was used.^{23,31} The homology model (mouse; UniProt P20917) is based on the three-dimensional structure of sialoadhesin³⁴ (mouse; Sig-



Scheme 2. Reagents and conditions: (a) NaOMe, MeOH, 94%; (b) (i) All-Br, 15-crown-5, DCM/50% NaOH, 60 °C; (ii) aq NaOH; (iii) TMS-CHN₂, toluene/MeOH, 79%; (c) (i) cat. OsO₄, NaIO₄, 2,6-lutidine, dioxane/H₂O; (ii) NaBH₄, MeOH, 0 °C, 65%; (d) (i) H₂ (1 atm), Pd/C, dioxane/NEt₃ (10:1); (ii) Ac₂O/py, 64%; (e) NaOMe, MeOH, 86%; (f) TsCl, NEt₃, cat. DMAP, DCM, 0 °C to rt, 84%; (g) NaN₃, 15-crown-5, DMF, 60 °C, 83%; (h) (i) H₂ (1 atm), cat. Pd/C, MeOH/NEt₃ (10:1); (ii) *p*-ClBzCl, cat. DMAP, NEt₃, DCM, 0 °C to rt, 66%; (i) (i) TFA/DCM (1:2), 0 °C to rt; (ii) FAcCl, cat. DMAP, NEt₃, DCM, 0 °C to rt, 69%; (j) (i) LiOH, THF/H₂O (9:1); (ii) Dowex 50X8 (Na⁺), 83%.

Table 1
Relative inhibitory concentrations (rIC₅₀) relative to reference compound **29**, K_D values of sialosides **2** and **29** and non-carbohydrate mimic **3**

Entry	Compound	rIC ₅₀ ^a	K _D (μM)
1		1	26
2		0.02	0.5
3		4300	>800

^a rIC₅₀ is the concentration when 50% of the protein are inhibited, measured relative to reference compound **29**.

lec-1, PDB code 1QFP, 2.8 Å resolution) and was generated following standard protocols using the Fugue⁴² and Orchestral⁴³ concepts as implemented in SYBYL 7.3.⁴⁴ The homology between sialoadhesin and MAG was calculated using the Clustal W multiple sequence alignment program.⁴⁵ The primary sequences of the two proteins and the residues of their binding sites exhibit a homology of 52.5% and 50.0%, respectively. After generation, the homology model was subjected to a full refinement in aqueous solution using AMBER 7.0.⁴⁶

The three-dimensional structures of compounds **2** and **3** were generated using MacroModel,⁴⁷ followed by a conformational search to identify low-energy conformations suited for binding (MacroModel,⁴⁷ OPLS force field,⁴⁸ aqueous solution, 5000 sampled structures). CM1 atomic partial charges were then generated using AMSOL.⁴⁹ In the main step, the ligands were manually docked to the binding site of the MAG model using the salt bridge to Arg118 as anchor point (Fig. 3). After rigorous molecular-dynamics relaxation (cf. below), each protein–ligand complex was fully minimized in aqueous solution.

A close inspection of the docking modes reveals slightly different binding modes for compounds **2** and **3** (Fig. 3). Beside the anchor point which was used for the docking of **2** and **3**, the hydrogens of the 5-NHAc groups interact with the backbone carbonyl of Gln126, and the C(4)-OH (in **2**) or C(4)-NH (in **3**), respectively, are engaged in hydrogen bonds with the side chain carbonyl of Gln126. In addition, the 2,3-difluorobenzyl aglycone in **2** as well as the benzyloxy substituent in **3** are hosted by a hydrophobic pocket formed by Trp59, Tyr60, Tyr69, and Tyr116. However, whereas in **2** the C(9)-NH of the benzamido group establishes a H-bond with the backbone carbonyl of Thr128, in **3** the 9-benzamido carbonyl interacts with the backbone NH of Thr128, thus forming a much weaker H-bond than the parent compound **2**. Consequently, the orientations of the two side chains, the glycerol side chain in **2** and the 2-benzamido-ethyl side chain in **3**, are slightly modified, leading to a noticeably reduced interaction of the *p*-chlorobenzamide in **3** with the second hydrophobic pocket formed by Ser130 and Glu131.

Since the above discussed docking modes are the result of thermodynamic considerations, we also performed extended molecular-dynamics simulations to assess the dynamic stability of the protein–ligand complexes using Desmond⁵¹ with an equilibrated water box (5458 and 4838 water molecules, respectively) and periodic boundary conditions. After an equilibration phase of 72 × 10–12 s, the structures were sampled for 4 × 10^{–9} s at 300 K and the essential interactions involved in the thermodynamic stabilization of the MAG–compound **2** complex were analyzed in time. An analogous study was done for the MAG–compound **3** complex. The results are summarized in Table 2.

The two complexes are characterized by obvious differences in respect to the hydrogen-bond network. While frequency and strength of the salt bridge to Arg118 is comparable for both **2** and **3** (entries 1 and 2), the hydrogen bonds of the 5-fluoroacetamido NH with the backbone carbonyl of Gln126 are comparable in frequency, but the interaction with **2** is slightly stronger (average distance 1.88 Å for **2** vs 1.93 Å for **3**, entry 3).

However, the major difference originates from the interaction patterns of the side chains, that is, the glycerol side chain in **2** and the 2-benzamidoethoxy side chain in **3**. In **2**, two important hydrogen bonds contribute to binding: (i) the 9-benzamido NH interacts with the backbone carbonyl of Thr128 (1.96 Å, entry 4) and (ii) the backbone NH of the same amino acid acts as donor in a H-bond with the oxygen of the C(8)-OH (1.94 Å, entry 6).

Since the benzamidoethoxy side chain in **3** is oriented in an inverted manner, a significantly weaker H-bond between the backbone NH of Thr128 and the 9-benzamido carbonyl (2.10 Å, frequency 0.1, entry 5) is formed. In addition, the hydroxy group at C(8) is absent, therefore the second H-bond is not present at all. As a consequence, the orientation of the side chain alters leading to a markedly reduced hydrophobic contact of the *p*-chlorobenzamide with Ser130 and Glu131.

Finally, the hydrogen bond of the C(4)-OH in **2** with the side chain carbonyl of Gln126, which was identified in the static docking studies, turned out to be irrelevant in the dynamic considerations. Although the hydrogen bond of the 4-acetamido NH in **3**

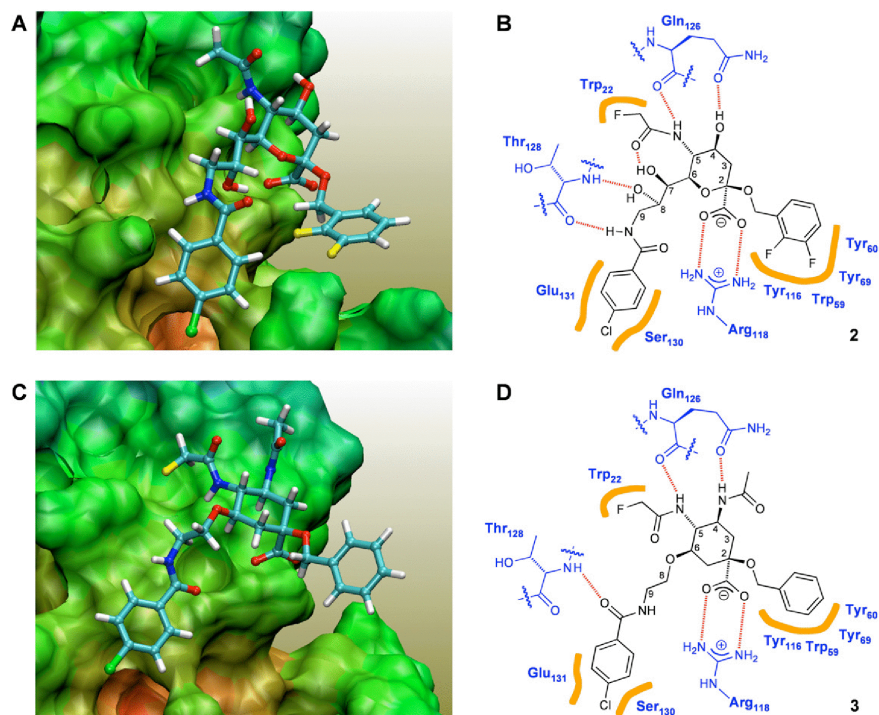


Figure 3. (A and C) The homology model of MAG complexed with **2** (A) and **3** (C). The images have been generated using VMD.⁵⁰ The ligands are depicted colored by atom (C: cyan, H: white, O: red, N: blue, Cl: green, F: yellow), the protein is represented by its van der Waals surface colored according to the depth. (B and D) Binding interactions of **2** (B) and **3** (D); hydrogen bonds are indicated as red dashed lines, hydrophobic interactions are shown as orange bars. For better comparability, the carbon atoms in **3** are numbered in analogy to compound **2**.

with the side chain carbonyl of Gln126 substantially contributes to the binding to MAG (1.93 Å, entry 7), this additional interaction unexpectedly does not compensate the experimentally observed loss of affinity caused by the reduced hydrophobic contact of the benzamido group in **3** compared to **2**.

3. Conclusion

The non-carbohydrate mimic **3** of the high-affinity MAG antagonist **2** was synthesized with the goal to improve its pharmacokinetic properties. Although the mimic **3** can establish all but one of the essential hydrogen bond interactions present in **2**, a 200,000-fold drop in affinity was observed. The docking modes of **2** and **3** to a homology model of MAG revealed a different orientation of the side chains in the 6-position. Obviously, the C(8)-OH and the C(9)-NH in sialoside **2** represent key polar groups,⁵² that is, the recognition of both of these hydrogen bond donors is required for binding. Consequently, in the absence of the stabilizing H-bond formed between the C(8)-OH in **2** and the backbone NH of Thr128, the benzamidoethoxy side chain adopts an inverted orientation, leading to a less favorable interaction of its benzamido group with the hydrophobic pocket formed by Ser130 and Glu131 and therefore to the experimentally observed dramatic loss in affinity.

In summary, while docking studies based on a thermodynamic inspection of the binding event do only insufficiently depict the MAG-mimetic **3** interaction, dynamic considerations based on

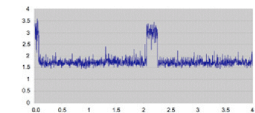
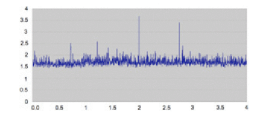
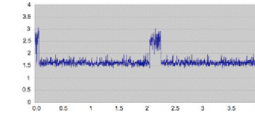
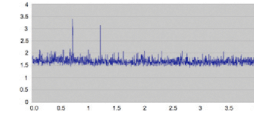
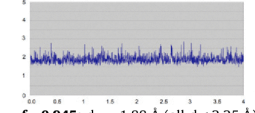
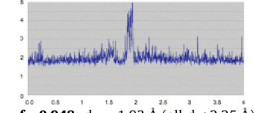
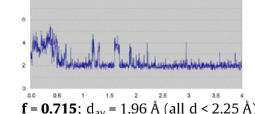
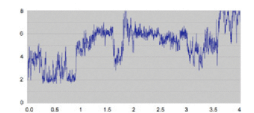
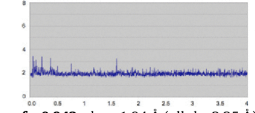

molecular-dynamics simulations gave a more detailed picture and allowed to interpret the experimental findings.

4. Experimental part

4.1. Chemistry

NMR spectra were recorded on a Bruker Avance DMX-500 (500 MHz) spectrometer. Assignment of ¹H and ¹³C NMR spectra was achieved using 2D methods (COSY, HSQC). Chemical shifts are expressed in ppm using residual CHCl₃ and H₂O as references. Optical rotations were measured using a Perkin-Elmer Polarimeter 341. Electron spray ionization mass spectra (ESI-MS) were obtained on a Waters micromass ZQ. The HRMS analyses were carried out using a Bruker QTOF. Reactions were monitored by TLC using glass plates coated with silica gel 60 F₂₅₄ (Merck) and visualized by using UV light and/or by charring with a molybdate solution (a 0.02 M solution of ammonium cerium sulfate dihydrate and ammonium molybdate tetrahydrate in aqueous 10% H₂SO₄). Column chromatography was performed on silica gel (Fluka, 40–60 mesh). MPLC separations were carried out on a CombiFlash Companion from Teledyne Isco equipped with RediSep normal-phase flash columns. Tetrahydrofuran (THF) and dioxane were freshly distilled under argon over sodium and benzophenone. Methanol (MeOH) was dried by refluxing with sodium methoxide and distilled immediately before use. Pyridine was freshly distilled under argon over CaH₂. Acetone, dichloromethane (DCM), *N,N*-

Table 2
Time-resolved H-bond distances between key amino-acid residues of MAG and compounds **2** and **3**

Entry	Interaction	Compound 2	Compound 3
1	$-\text{COO}^{\ominus}-\text{H}_2\text{N}^{\oplus}-\text{Arg118}$ (pair #1)	 $f = 0.922$; $d_{av} = 1.71 \text{ \AA}$ (all $d < 2.0 \text{ \AA}$)	 $f = 0.973$; $d_{av} = 1.69 \text{ \AA}$ (all $d < 2.0 \text{ \AA}$)
2	$-\text{COO}^{\ominus}-\text{H}_2\text{N}^{\oplus}-\text{Arg118}$ (pair #2)	 $f = 0.934$; $d_{av} = 1.62 \text{ \AA}$ (all $d < 2.0 \text{ \AA}$)	 $f = 0.984$; $d_{av} = 1.66 \text{ \AA}$ (all $d < 2.0 \text{ \AA}$)
3	5-NH-O=C-Gln126 _{MC}	 $f = 0.945$; $d_{av} = 1.88 \text{ \AA}$ (all $d < 2.25 \text{ \AA}$)	 $f = 0.948$; $d_{av} = 1.93 \text{ \AA}$ (all $d < 2.25 \text{ \AA}$)
4	9-NH-O=C-Thr128	 $f = 0.715$; $d_{av} = 1.96 \text{ \AA}$ (all $d < 2.25 \text{ \AA}$)	absent
5	<i>p</i> -ClPh-CO-HN-Thr128	absent	 $f = 0.102$; $d_{av} = 2.10 \text{ \AA}$ (all $d < 2.25 \text{ \AA}$)
6	8-O-HN-Thr128 H	 $f = 0.942$; $d_{av} = 1.94 \text{ \AA}$ (all $d < 2.25 \text{ \AA}$)	absent
7	4-NH-O=C-Gln126 _{SC}	absent	 $f = 0.693$; $d_{av} = 1.93 \text{ \AA}$ (all $d < 2.25 \text{ \AA}$)

The horizontal axis represents the simulation time, the vertical axis the H...acceptor separation. *f*: frequency of observance of the H-bond throughout the molecular-dynamics simulation; d_{av} : average H...acceptor distance; MC: main chain; SC: side chain.

dimethylformamide (DMF), and toluene were dried by filtration over Al_2O_3 (Fluka, type 5016 A basic).

4.1.1. 3,4-O-Isopropylidene-1,5-quinic lactone (**5**)³⁷

To a solution of (–)-quinic acid (**4**, 15.0 g, 78.1 mmol) and DMP (33.5 mL, 273 mmol) in acetone (75 mL) was added *p*-toluene sulfonic acid monohydrate (149 mg, 0.78 mmol). The reaction mixture was stirred at 70 °C for 3 h. After cooling to rt, NEt_3 (3 mL) was added with stirring and the solvent was evaporated. The residue was purified by chromatography on silica gel (petroleum ether/EtOAc 3:1 to 1:1, +0.5% NEt_3) to yield **5** (14.7 g, 87%) as a colorless solid. The analytical data correspond to reported values.³⁷

4.1.2. 1-O-Benzyl-3,4-O-isopropylidene-1,5-quinic lactone (**6**)

NaH (3.53 g, 60% suspension in oil, 88.3 mmol) was washed with *n*-hexane ($2 \times 5 \text{ mL}$) under argon and suspended in DMF (10 mL) at 0 °C. Then a solution of **5** (9.46 g, 44.2 mmol) in DMF

(50 mL) was added dropwise during 30 min with stirring. The reaction mixture was warmed to rt, stirred for 19 h and then quenched by slowly adding saturated aqueous NH_4Cl (20 mL) at 0 °C. The mixture was extracted with DCM ($3 \times 40 \text{ mL}$) and the combined organic layers were dried with Na_2SO_4 . After filtration and evaporation of the solvents, the residue was purified by chromatography on silica (petroleum ether/EtOAc 10:1 to 5:1, +0.5% NEt_3) to yield **6** (12.6 g, 94%) as a colorless solid.

^1H NMR (500 MHz, CDCl_3): δ 1.34, 1.52 (2 s, 6H, $\text{C}(\text{CH}_3)_2$), 2.26 (dd, $J = 2.7, 15.0 \text{ Hz}$, 1H, H-2a), 2.46–2.56 (m, 3H, H-2b, H-6), 4.32 (dt, $J = 1.2, 6.4 \text{ Hz}$, H-4), 4.55 (dt, $J = 2.7, 7.0 \text{ Hz}$, H-3), 4.58, 4.67 (A, B of AB, $J = 10.8 \text{ Hz}$, 2H, CH_2Ph), 4.70 (dd, $J = 2.4, 6.0 \text{ Hz}$, 1H, H-5), 7.29–7.40 (m, 5H, C_6H_5); ^{13}C NMR (125 MHz, CDCl_3): δ 24.3, 27.0 ($\text{C}(\text{CH}_3)_2$), 30.7 (C-6), 36.1 (C-2), 67.0 (CH_2Ph), 71.4 (C-3), 72.4 (C-4), 75.0 (C-5), 76.9 (C-1), 109.6 ($\text{C}(\text{CH}_3)_2$), 127.8, 127.9, 128.4, 137.4 (6C, C_6H_5), 175.8 (CO); ESI-MS Calcd for $\text{C}_{17}\text{H}_{20}\text{NaO}_5$ [$\text{M}+\text{Na}$] $^+$: 327.12. Found: 327.08.

4.1.3. Methyl (1R,3R,4S,5R)-1-benzyloxy-3,4,5-trihydroxy-4,5-O-isopropylidene-cyclohexanecarboxylate (7)

To a suspension of **6** (9.84 g, 32.3 mmol) in MeOH (170 mL) was added a freshly prepared solution of sodium (820 mg, 35.6 mmol) in MeOH (50 mL) at 0 °C under argon. The reaction mixture was stirred for 3 h at rt, then the solution was neutralized by adding Dowex 50X8 ion-exchange resin. After filtration and evaporation of the solvents, the residue was purified by chromatography on silica (petroleum ether/EtOAc 2:1 to 1:8, +0.5% NEt₃) to yield **7** (7.96 g, 73%) as a colorless solid.

¹H NMR (500 MHz, CDCl₃): δ 1.37, 1.48 (2 s, 6H, C(CH₃)₂), 1.81 (dd, *J* = 10.9, 13.6 Hz, 1H, H-6a), 2.25–2.31 (m, 2H, H-2a, H-6b), 2.54 (m, 1H, H-2b), 3.78 (m, 3H, OMe), 3.94 (t, *J* = 6.4 Hz, H-4), 4.16 (ddd, *J* = 4.1, 7.0, 10.9 Hz, 1H, H-5), 4.34 (A of AB, *J* = 10.2 Hz, 1H, CH₂Ph), 4.46 (dt, *J* = 3.6, 5.3 Hz, H-3), 4.60 (B of AB, *J* = 10.2 Hz, 1H, CH₂Ph), 7.27–7.39 (m, 5H, C₆H₅); ¹³C NMR (125 MHz, CDCl₃): δ 25.9, 28.2 (C(CH₃)₂), 30.9 (C-2), 37.7 (C-6), 52.5 (OMe), 67.1 (CH₂Ph), 68.1 (C-5), 72.9 (C-3), 79.2 (C-1), 80.1 (C-4), 109.1 (C(CH₃)₂), 127.9, 128.2, 128.3, 138.0 (6C, C₆H₅), 174.1 (CO); ESI-MS Calcd for C₁₈H₂₅O₆ [M+H]⁺: 337.17. Found: 337.14.

4.1.4. Methyl (1S,3R,4R,5R)-1-benzyloxy-3,4-dihydroxy-3,4-O-isopropylidene-5-(*p*-tosyloxy)-cyclohexanecarboxylate (8)

To a solution of **7** (5.55 g, 16.5 mmol) and DMAP (190 mg) in pyridine (40 mL) was added *p*-tosyl chloride (6.29 g, 33.0 mmol) under argon. The solution was stirred for 19 h, then the pyridine was evaporated and water (50 mL) was added. The aqueous phase was extracted with DCM (3 × 50 mL), then the combined organic layers were washed with brine (20 mL) and dried with Na₂SO₄. After evaporation of the solvent, the residue was purified by chromatography on silica (petroleum ether/EtOAc 3:1 to 1:1, +0.5% NEt₃) to give **8** (7.96 g, 98%) as a yellowish oil.

¹H NMR (500 MHz, CDCl₃): δ 0.93, 1.23 (2 s, 6H, C(CH₃)₂), 1.91 (t, *J* = 12.7 Hz, 1H, H-6a), 2.16 (dd, *J* = 5.0, 16.1 Hz, 1H, H-2a), 2.39 (s, 3H, PhCH₃), 2.59 (d, *J* = 16.1 Hz, 1H, H-2b), 2.68 (dd, *J* = 3.1, 13.6 Hz, 1H, H-6b), 3.76 (m, 3H, OMe), 4.01 (t, *J* = 6.5 Hz, H-4), 4.40–4.43 (m, 2H, H-3, CH₂Ph), 4.55 (B of AB, *J* = 10.5 Hz, 1H, CH₂Ph), 4.64 (m, 1H, H-5), 7.19 (AA' of AA'BB', *J* = 8.0 Hz, 2H, C₆H₄), 7.31–7.37 (m, 5H, C₆H₅), 7.71 (BB' of AA'BB', *J* = 8.0 Hz, 2H, C₆H₄); ¹³C NMR (125 MHz, CDCl₃): δ 21.5 (PhCH₃), 25.8, 27.3 (C(CH₃)₂), 31.1 (C-2), 35.5 (C-6), 52.5 (OMe), 66.7 (CH₂Ph), 73.2 (C-3), 76.6 (C-4), 79.1, 79.2 (C-1, C-5), 109.2 (C(CH₃)₂), 127.4, 127.5, 128.1, 128.4, 129.6, 132.9, 137.9 (12C, C₆H₄, C₆H₅), 172.6 (CO); ESI-MS Calcd for C₂₅H₃₀NaO₈S [M+Na]⁺: 513.16. Found: 513.18.

4.1.5. Methyl (1S,3R,4R,5R)-1-benzyloxy-3,4-dihydroxy-5-(*p*-tosyloxy)-cyclohexanecarboxylate (9)

A suspension of **8** (7.96 g, 16.2 mmol) in 80% aqueous acetic acid (30 mL) was stirred for 2.5 h at 60 °C. After cooling to rt, the solvents were removed by co-evaporation with toluene (3 × 50 mL). The residue was dried in high vacuum to give **9** (7.13 g, 98%), which was used without further purification.

¹H NMR (500 MHz, CDCl₃): δ 1.94 (dd, *J* = 3.1, 15.1 Hz, 1H, H-2a), 2.16 (dd, *J* = 11.9, 14.2 Hz, 1H, H-6a), 2.40 (m, 1H, H-2b), 2.43 (s, 3H, PhCH₃), 2.89 (dt, *J* = 3.4, 14.3 Hz, 1H, H-6b), 3.56 (dd, *J* = 3.5, 9.3 Hz, H-4), 3.78 (m, 3H, OMe), 4.05 (m, 1H, H-3), 4.50, 4.52 (A, B of AB, *J* = 10.2 Hz, 2H, CH₂Ph), 4.59 (m, 1H, H-5), 7.31 (AA' of AA'BB', *J* = 8.1 Hz, 2H, C₆H₄), 7.33–7.39 (m, 5H, C₆H₅), 7.82 (BB' of AA'BB', *J* = 8.1 Hz, 2H, C₆H₄); ¹³C NMR (125 MHz, CDCl₃): δ 21.7 (PhCH₃), 34.5 (C-6), 37.3 (C-2), 52.7 (OMe), 67.6 (CH₂Ph), 70.0 (C-3), 72.9 (C-4), 78.5 (C-5), 81.5 (C-1), 128.1, 128.3, 128.7, 129.8, 133.2, 136.5, 145.0 (12C, C₆H₄, C₆H₅), 171.1 (CO); ESI-MS Calcd for C₂₂H₂₆NaO₈S [M+Na]⁺: 473.12. Found: 473.15.

4.1.6. Methyl (1R,3S,4S,5R)-1-benzyloxy-3,4-epoxy-5-hydroxy-cyclohexane-1-carboxylate (10)

To a solution of **9** (7.13 g, 15.8 mmol) in THF (50 mL) was added DBU (2.39 mL, 15.8 mmol) dropwise during 10 min at 0 °C under argon. The mixture was stirred at 0 °C for 1 h, then warmed to rt and stirred for another 17 h. The solvent was removed, and the residue was purified by MPLC on silica (petroleum ether/EtOAc) to yield **10** (4.23 g, 96%) as a colorless solid.

¹H NMR (500 MHz, CDCl₃): δ 1.86 (dd, *J* = 9.5, 12.6 Hz, 1H, H-6a), 2.13 (d, *J* = 15.1 Hz, 1H, H-2a), 2.23 (m, 1H, H-6b), 2.68 (ddd, *J* = 2.0, 4.8, 15.1 Hz, 1H, H-2b), 3.35 (m, 1H, H-4), 3.44 (t, *J* = 4.4 Hz, 1H, H-3), 3.77 (s, 3H, OMe), 4.03 (m, 1H, H-5), 4.31, 4.49 (A, B of AB, *J* = 10.6 Hz, 2H, CH₂Ph), 7.27–7.38 (m, 5H, C₆H₅); ¹³C NMR (125 MHz, CDCl₃): δ 30.3 (C-2), 36.4 (C-6), 52.5 (C-3), 52.8 (OMe), 54.8 (C-4), 65.9 (C-5), 66.9 (CH₂Ph), 79.5 (C-1), 127.7, 127.8, 128.4, 137.4 (6C, C₆H₅), 173.2 (CO); ESI-MS Calcd for C₁₅H₁₈NaO₅ [M+Na]⁺: 301.11. Found: 301.14.

4.1.7. Methyl (1R,3S,4S,5R)-5-benzoyloxy-1-benzyloxy-3,4-epoxy-cyclohexane-1-carboxylate (11)

To a solution of **10** (2.10 g, 7.55 mmol) in pyridine (20 mL) were subsequently added benzoyl chloride (1.78 mL, 15.1 mmol) and DMAP (120 mg) at 0 °C under argon. The mixture was stirred for 4 h at rt and then concentrated in vacuo. The residue was co-evaporated with toluene (3 × 10 mL), dissolved in DCM (20 mL), and washed with 1 N HCl (10 mL) and saturated aqueous NaHCO₃ (10 mL). The organic phase was dried (Na₂SO₄), filtered and concentrated in vacuo. The residue was purified by MPLC on silica (petroleum ether/EtOAc) to afford **11** (2.79 g, 97%) as a colorless solid.

[α]_D –33.4 (c 0.50, CHCl₃); ¹H NMR (500 MHz, CDCl₃): δ 2.06–2.15 (m, 2H, H-2a, H-6a), 2.42 (dd, *J* = 4.3, 11.9 Hz, 1H, H-6b), 2.85 (dd, *J* = 3.2, 14.9 Hz, 1H, H-2b), 3.43 (t, *J* = 4.5 Hz, 1H, H-3), 3.47 (m, 1H, H-4), 3.84 (s, 3H, OMe), 4.35, 4.50 (A, B of AB, *J* = 10.9 Hz, 2H, CH₂Ph), 5.39 (m, 1H, H-5), 7.28–7.35, 7.43–7.46, 7.56–7.59, 8.07–8.09 (m, 10H, 2 C₆H₅); ¹³C NMR (125 MHz, CDCl₃): δ 31.2 (C-2), 32.2 (C-6), 51.8 (C-3), 52.7, 52.8 (C-4, OMe), 66.8 (CH₂Ph), 69.2 (C-5), 79.3 (C-1), 127.6, 127.8, 128.4, 128.4, 129.7, 129.8, 133.3, 137.5 (12C, 2C₆H₅), 166.0 (COPh), 173.2 (CO₂Me); ESI-MS Calcd for C₂₂H₂₂NaO₆ [M+Na]⁺: 405.13. Found: 405.08.

4.1.8. Methyl (1R,3R,4S,5R)-5-azido-3-benzoyloxy-1-benzyloxy-4-hydroxy-cyclohexane-1-carboxylate (12)

To a solution of **11** (2.79 g, 7.30 mmol) in MeOH/water (8:1, 54 mL) were added sodium azide (2.47 g, 37.9 mmol) and ammonium chloride (859 mg, 16.1 mmol), and the mixture was stirred at 70 °C for 18 h under argon. The reaction mixture was diluted with saturated aqueous NaHCO₃ (10 mL) and water (20 mL), and the solution was carefully evaporated to remove methanol. The resulting aqueous phase containing an oily residue was extracted with ethyl acetate (3 × 50 mL). The combined organic extracts were dried (Na₂SO₄), filtered, and concentrated in vacuo. The residue was purified by MPLC on silica (1% gradient of MeOH in DCM) to yield **12** (2.45 g, 78%) as a colorless oil.

[α]_D –53.6 (c 0.45, CHCl₃); IR (film) 3479 (m b, OH), 2107 (vs, N₃), 1722 (vs, CO) cm⁻¹; ¹H NMR (500 MHz, CDCl₃): δ 1.87 (dd, *J* = 11.7, 13.8 Hz, 1H, H-6a), 2.23 (dd, *J* = 3.5, 15.9 Hz, 1H, H-2a), 2.44 (d, *J* = 5.1 Hz, 1H, OH), 2.59 (m, 1H, H-6b), 2.67 (dt, *J* = 3.2, 15.9 Hz, 1H, H-2b), 3.76–3.80 (m, 4H, H-4, OMe), 4.10 (m, 1H, H-5), 4.31, 4.38 (A, B of AB, *J* = 10.5 Hz, 2H, CH₂Ph), 5.65 (q, *J* = 3.5 Hz, 1H, H-3), 7.20–7.28, 7.48–7.53, 7.90–7.93 (m, 10H, 2C₆H₅); ¹³C NMR (125 MHz, CDCl₃): δ 32.3 (C-2), 36.8 (C-6), 52.6 (OMe), 58.4 (C-5), 67.4 (CH₂Ph), 70.8 (C-3), 73.7 (C-4), 79.4 (C-1), 127.7, 127.8, 128.3, 128.4, 129.4, 129.9, 133.3, 137.3 (12C, 2C₆H₅), 166.2 (COPh), 172.7 (CO₂Me); ESI-MS Calcd for C₂₂H₂₃N₃NaO₆ [M+Na]⁺: 448.15. Found: 448.16.

4.1.9. Methyl (1R,3R,4S,5R)-5-azido-3-benzoyloxy-1-benzoyloxy-4-(methylsulfonyloxy)-cyclohexane-1-carboxylate-1,4-lactam (13)

To a solution of **12** (2.43 g, 5.70 mmol) in DCM (25 mL) at 0 °C under argon was added NET_3 (3.18 mL, 22.8 mmol) followed by the addition of methanesulfonyl chloride (887 μL , 11.4 mmol). The mixture was stirred at 0 °C for 1 h and warmed to rt with stirring for 3 h. The reaction was quenched with MeOH (2 mL) and concentrated in vacuo. Purification of the residue by MPLC on silica (petrol ether/EtOAc) yielded **13** (2.69 g, 94%) as a yellowish oil.

$[\alpha]_{\text{D}}^{25} = -31.0$ (c 0.50, CHCl_3); IR (film) 2111 (vs, N_3), 1732 (vs, CO) cm^{-1} ; ^1H NMR (500 MHz, CDCl_3): δ 2.01 (m, 1H, H-6a), 2.35 (d, $J = 15.9$ Hz, 1H, H-2a), 2.57–2.73 (m, 2H, H-2b, H-6b), 3.08 (s, 3H, SMe), 3.79 (s, 3H, OMe), 4.27–4.34 (m, 2H, H-5, $\text{CH}_2\text{Ph-H}_A$), 4.38 (B of AB, $J = 10.5$ Hz, 1H, $\text{CH}_2\text{Ph-H}_B$), 4.70 (dd, $J = 3.4$, 9.6 Hz, 1H, H-4), 5.84 (m, 1H, H-3), 7.18–7.32, 7.53–7.55, 7.94–7.95 (m, 10H, $2\text{C}_6\text{H}_5$); ^{13}C NMR (125 MHz, CDCl_3): δ 31.8 (C-2), 37.4 (C-6), 38.7 (SMe), 52.7 (OMe), 55.9 (C-5), 67.4 (CH_2Ph), 68.5 (C-3), 79.8 (2C, C-1, C-4), 127.8, 127.9, 128.3, 128.5, 129.4, 129.9, 133.4, 137.0 (12C, $2\text{C}_6\text{H}_5$), 166.3 (COPh), 172.5 (CO_2Me); ESI-MS Calcd for $\text{C}_{23}\text{H}_{25}\text{N}_3\text{NaO}_8\text{S}$ $[\text{M}+\text{Na}]^+$: 526.13. Found: 526.09.

4.1.10. Methyl (1R,3R,4R,5R)-4,5-aziridino-3-benzoyloxy-1-benzoyloxy-cyclohexane-1-carboxylate (14)

To a solution of **13** (2.69 g, 5.34 mmol) in THF (30 mL) at 0 °C was added PPh_3 (1.76 g, 6.68 mmol), initially adding a third of the amount while cooling and then after removing the ice bath adding the remaining amount over a period of 15 min. The reaction mixture was stirred at rt for 4 h, then NET_3 (1.12 mL, 8.02 mmol) and water (2 mL) were added, and the mixture was stirred at rt for 16 h. The reaction mixture was concentrated to remove THF, and the residue was partitioned between DCM (50 mL) and brine (20 mL). The aqueous phase was extracted with DCM (3×20 mL), and the combined organic extracts were dried (Na_2SO_4), filtered, and evaporated. The residue was purified by MPLC on silica (petrol ether/EtOAc, +1% NET_3) to give 2.20 g of a 3:2-mixture of **14** and triphenylphosphine oxide, which was directly used in the next step.

^1H NMR (500 MHz, CDCl_3): δ 2.13 (dd, $J = 3.7$, 14.8 Hz, 1H, H-2a), 2.37–2.44 (m, 3H, H-2b, H-6), 2.45–2.50 (m, 2H, H-4, H-5), 3.74 (s, 3H, OMe), 4.38, 4.53 (A, B of AB, $J = 11.1$ Hz, 1H, CH_2Ph), 5.60 (t, $J = 4.9$ Hz, 1H, H-3), 7.25–7.33, 7.53–7.56, 7.96–7.97 (m, 10H, $2\text{C}_6\text{H}_5$); ^{13}C NMR (125 MHz, CDCl_3): δ 28.1 (C-6), 30.8 (C-2), 31.6, 32.5 (C-4, C-5), 52.2 (OMe), 66.5 (CH_2Ph), 68.1 (C-3), 76.9 (C-1), 127.5, 128.2, 128.3, 128.5, 129.7, 132.9, 138.2 (12C, $2\text{C}_6\text{H}_5$), 166.1 (COPh), 173.1 (CO_2Me); ESI-MS Calcd for $\text{C}_{22}\text{H}_{24}\text{NO}_5$ $[\text{M}+\text{H}]^+$: 382.17. Found: 382.10.

4.1.11. (1R,3R,4R,5S)-4-Amino-5-azido-1-benzoyloxy-3-benzoyloxy-cyclohexane-1-carboxylate-1,4-lactam (16)

To a solution of **14** (105 mg, 0.275 mmol) in DMF (3 mL) was added NaN_3 (89.3 mg, 1.37 mmol) and NH_4Cl (29.4 mg, 0.55 mmol), and the mixture was heated at 65 °C under argon for 23 h. The reaction mixture was cooled to rt, diluted with ethyl acetate (10 mL), and filtered. The filtrate was washed with brine (5 mL) and the aqueous phase was extracted with EtOAc (3×10 mL). The combined organic phases were dried (Na_2SO_4), filtered, and concentrated in vacuo. The residue was purified by chromatography on silica (petrol ether/EtOAc 2:1 to 1:1) to yield **16** (53.5 mg, 50%) as colorless foam.

$[\alpha]_{\text{D}}^{25} = -29.6$ (c 1.24, CHCl_3); IR (film) 2099 (vs, N_3), 1719 (vs, CO) cm^{-1} ; ^1H NMR (500 MHz, CDCl_3): δ 1.99 (dd, $J = 5.5$, 13.4 Hz, 1H, H-6a), 2.04 (dd, $J = 3.8$, 13.7 Hz, 1H, H-2a), 2.57 (ddd, $J = 3.5$, 10.8, 13.3 Hz, 1H, H-6b), 2.81 (ddd, $J = 3.5$, 10.5, 13.7 Hz, 1H, H-2b), 4.04 (m, 1H, H-4), 4.09 (m, 1H, H-5), 4.78, 4.84 (A, B of AB, $J = 11.0$ Hz, 2H, CH_2Ph), 5.35 (m, 1H, H-3), 7.28–7.36, 7.55–7.61,

8.09–8.10 (m, 10H, $2\text{C}_6\text{H}_5$); ^{13}C NMR (125 MHz, CDCl_3): δ 34.8 (C-6), 36.0 (C-2), 49.8 (C-4), 57.9 (C-5), 67.1 (CH_2Ph), 70.1 (C-3), 74.1 (C-1), 127.7, 128.4, 128.5, 129.3, 129.8, 133.4, 138.2 (12C, $2\text{C}_6\text{H}_5$), 166.1, 175.2 (2 CO); ESI-MS Calcd for $\text{C}_{21}\text{H}_{20}\text{N}_4\text{NaO}_4$ $[\text{M}+\text{Na}]^+$: 415.14. Found: 415.16.

4.1.12. Methyl (1R,3R,4R,5R)-4,5-aziridino-3-benzoyloxy-1-benzoyloxy-N-(tert-butoxycarbonyl)-cyclohexane-1-carboxylate (17)

Crude aziridine **14** (2.10 g, 3:2 mixture with PPh_3O) was dissolved in DCM (50 mL) and cooled to 0 °C under argon. Then NET_3 (1.53 mL, 11.0 mmol), Boc_2O (1.80 g, 8.25 mmol), and DMAP (10 mg) were subsequently added and the solution was stirred at rt for 5 h. The solvent was evaporated and the residue purified by MPLC on silica (1% gradient of EtOAc in petrol ether) to yield **17** (1.66 g, 96%) as a colorless oil.

^1H NMR (500 MHz, CDCl_3): δ 1.47 (s, 9H, $\text{C}(\text{CH}_3)_3$), 2.26 (d, $J = 15.3$ Hz, 1H, H-2a), 2.35 (dd, $J = 5.4$, 15.4 Hz, 1H, H-2b), 2.40 (d, $J = 15.3$ Hz, 1H, H-6a), 2.52 (ddd, $J = 1.6$, 5.8, 15.4 Hz, 1H, H-6b), 2.83 (dd, $J = 4.9$, 5.7 Hz, 1H, H-5), 2.87 (m, 1H, H-4), 3.74 (s, 3H, OMe), 4.31, 4.50 (A, B of AB, $J = 11.0$ Hz, 2H, PhCH_2), 5.68 (m, 1H, H-3), 7.27–7.32, 7.50–7.53, 7.93–7.95 (m, 10H, $2\text{C}_6\text{H}_5$); ^{13}C NMR (125 MHz, CDCl_3): δ 27.8 ($\text{C}(\text{CH}_3)_3$), 30.3 (C-2), 31.5 (C-6), 34.8 (C-5), 38.1 (C-4), 52.3 (OMe), 66.6 (C-3), 66.7 (CH_2Ph), 76.5 (C-1), 81.7 ($\text{C}(\text{CH}_3)_3$), 127.4, 127.5, 128.2, 128.3, 129.8, 129.8, 133.0, 137.9 (12C, $2\text{C}_6\text{H}_5$), 161.6 (NCO), 165.8, 172.9 (2CO); ESI-MS Calcd for $\text{C}_{27}\text{H}_{31}\text{NNaO}_7$ $[\text{M}+\text{Na}]^+$: 504.20. Found: 504.19.

4.1.13. Methyl (1R,3R,4R,5S)-5-azido-1-benzoyloxy-3-benzoyloxy-4-(tert-butoxycarbonylamino)-cyclohexane-1-carboxylate (18) and methyl (1R,3R,4S,5R)-4-azido-1-benzoyloxy-3-benzoyloxy-5-(tert-butoxycarbonylamino)-cyclohexane-1-carboxylate (19)

To a solution of **17** (1.38 g, 2.87 mmol) in DMF (25 mL) was added sodium azide (1.86 g, 28.7 mmol) and ammonium chloride (1.23 g, 22.9 mmol), and the mixture was heated at 70 °C under argon for 18 h. The reaction mixture was cooled to rt, diluted with ethyl acetate (25 mL), and filtered. The filtrate was washed with brine (20 mL) and the aqueous phase was extracted with EtOAc (3×25 mL). The combined organic phases were dried (Na_2SO_4), filtered, and concentrated in vacuo. The residue was purified by MPLC on silica (petrol ether/EtOAc) to yield **18** (956 mg, 64%) and **19** (398 mg, 27%) as colorless foams.

Data for **18**: $[\alpha]_{\text{D}}^{25} = -49.0$ (c 1.05, CHCl_3); IR (KBr) 3385 (m, NH), 2978 (m), 2100 (vs, N_3), 1722 (vs, CO), 1516 (m), 1454 (m), 1391 (m), 1301 (s), 1278 (vs), 1205 (m), 1157 (s), 1112 (s) cm^{-1} ; ^1H NMR (500 MHz, CDCl_3): δ 1.27 (s, 9H, $\text{C}(\text{CH}_3)_3$), 1.75 (t, $J = 12.7$ Hz, 1H, H-6a), 1.92 (t, $J = 12.2$ Hz, 1H, H-2a), 2.79 (d, $J = 12.7$ Hz, 1H, H-6b), 2.87 (d, $J = 11.7$ Hz, 1H, H-2b), 3.56 (m, 1H, H-5), 3.85 (s, 3H, OMe), 3.91 (q, $J = 10.4$ Hz, 1H, H-4), 4.42, 4.47 (A, B of AB, $J = 10.8$ Hz, 2H, PhCH_2), 4.51 (d, $J = 9.9$ Hz, 1H, NH), 5.09 (m, 1H, H-3), 7.27–7.36, 7.41–7.44, 7.55–7.58, 8.03–8.04 (m, 10H, $2\text{C}_6\text{H}_5$); ^{13}C NMR (125 MHz, CDCl_3): δ 28.1 ($\text{C}(\text{CH}_3)_3$), 38.1 (C-2), 38.2 (C-6), 52.9 (OMe), 58.0 (C-4), 58.9 (C-5), 67.1 (CH_2Ph), 70.1 (C-3), 77.9 (C-1), 80.0 ($\text{C}(\text{CH}_3)_3$), 127.6, 127.9, 128.3, 128.4, 129.5, 129.9, 133.3, 137.2 (12C, $2\text{C}_6\text{H}_5$), 155.5 (NCO), 166.2, 172.4 (2CO); ESI-MS Calcd for $\text{C}_{27}\text{H}_{32}\text{N}_4\text{NaO}_7$ $[\text{M}+\text{Na}]^+$: 547.22. Found: 547.29.

Data for **19**: ^1H NMR (500 MHz, CDCl_3): δ 1.47 (s, 9H, $\text{C}(\text{CH}_3)_3$), 2.17–2.26 (m, 2H, H-2a, H-6a), 2.56 (d, $J = 13.7$ Hz, 1H, H-6b), 2.63 (d, $J = 14.4$ Hz, 1H, H-2b), 3.77 (s, 3H, OMe), 3.82 (m, 1H, H-4), 4.22 (m, 1H, H-5), 4.27, 4.39 (A, B of AB, $J = 10.4$ Hz, 2H, PhCH_2), 5.01 (br s, 1H, NH), 5.63 (m, 1H, H-3), 7.17–7.23, 7.27–7.30, 7.49–7.52, 7.94–7.95 (m, 10H, $2\text{C}_6\text{H}_5$); ^{13}C NMR (125 MHz, CDCl_3): δ 28.4 ($\text{C}(\text{CH}_3)_3$), 33.1 (C-2), 36.4 (C-6), 46.8 (C-5), 52.7 (OMe), 62.8 (C-4), 67.2 (CH_2Ph), 69.9 (C-3), 79.5 (C-1), 80.1 ($\text{C}(\text{CH}_3)_3$), 127.6, 127.8,

128.1, 128.3, 129.3, 129.9, 133.2, 137.3 (12C, 2C₆H₅), 155.1 (NCO), 165.8, 173.2 (2CO); ESI-MS Calcd for C₂₇H₃₂N₄NaO₇ [M+Na]⁺: 547.22. Found: 547.28.

4.1.14. Methyl (1R,3R,4R,5S)-5-azido-1-benzyloxy-4-(tert-butoxycarbonylamino)-3-hydroxy-cyclohexane-1-carboxylate (20)

A solution of **18** (1.01 g, 1.93 mmol) in MeOH (15 mL) was treated with 1 M methanolic NaOMe (1.5 mL) under argon at rt for 3 h. The reaction mixture was neutralized with acetic acid and concentrated. The residue was purified by MPLC on silica (petrol ether/EtOAc) to give **20** (765 mg, 94%) as a colorless foam.

[α]_D +5.4 (c 1.01, CHCl₃); IR (KBr) 3268 (s, OH), 2982 (m), 2101 (vs, N₃), 1736 (vs, CO), 1675 (vs, CO), 1551 (vs), 1455 (m), 1393 (w), 1309 (vs), 1276 (m), 1199 (s), 1175 (s), 1126 (s) cm⁻¹; ¹H NMR (500 MHz, CDCl₃): δ 1.46 (s, 9H, C(CH₃)₃), 1.71 (t, J = 12.5 Hz, 1H, H-6a), 1.72 (t, J = 12.3 Hz, 1H, H-2a), 2.75 (d, J = 12.5 Hz, 1H, H-6b), 2.76 (d, J = 12.3 Hz, 1H, H-2b), 3.34 (q, J = 9.2 Hz, 1H, H-4), 3.44 (m, OH), 3.50 (m, 1H, H-5), 3.62 (m, 1H, H-3), 3.77 (s, 3H, OMe), 4.41, 4.46 (A, B of AB, J = 10.8 Hz, 2H, PhCH₂), 4.77 (m, 1H, NH), 7.27–7.35 (m, 5H, C₆H₅); ¹³C NMR (125 MHz, CDCl₃): δ 28.3 (C(CH₃)₃), 37.7 (C-6), 40.7 (C-2), 52.6 (OMe), 58.6 (C-5), 60.8 (C-4), 67.0 (CH₂Ph), 69.0 (C-3), 77.9 (C-1), 80.8 (C(CH₃)₃), 127.7, 127.9, 128.4, 137.3 (6C, C₆H₅), 156.8 (NCO), 172.5 (CO); HRMS Calcd for C₂₀H₂₈N₄NaO₆ [M+Na]⁺: 443.1901. Found: 443.1902.

4.1.15. Methyl (1R,3R,4R,5S)-5-azido-1-benzyloxy-4-(tert-butoxycarbonylamino)-3-allyloxy-cyclohexane-1-carboxylate (21)

A solution of **20** (748 mg, 1.78 mmol) and 15-crown-5 (35.2 μL, 0.178 mmol) in DCM (10 mL) was added to 50% aqueous NaOH (25 mL, w/v). Allyl bromide (770 μL, 8.90 mmol) was then added, and the resultant mixture was refluxed with vigorous stirring for 16 h. The reaction mixture was concentrated to remove DCM, then MeOH (10 mL) was added and the solution stirred for 30 min at rt. After neutralization with 4 N aqueous HCl, the aqueous solution was extracted with EtOAc (4 × 50 mL). The combined organic layers were dried (Na₂SO₄) and concentrated in vacuo. The residue was dissolved in toluene (20 mL) and MeOH (15 mL) and treated with TMS-CHN₂ (2.5 mL, 2 M solution in hexane). After evaporation to dryness, the residue was purified by MPLC on silica (petrol ether/EtOAc) to give **21** (644 mg, 79%) as a colorless oil.

[α]_D –8.7 (c 1.02, CHCl₃); IR (film) 3375 (s, NH), 3066 (w), 2977 (s), 2098 (vs, N₃), 1719 (vs, CO), 1522 (vs), 1455 (m), 1391 (m), 1366 (s), 1300 (s), 1250 (s), 1203 (s), 1159 (vs) cm⁻¹; ¹H NMR (500 MHz, CDCl₃): δ 1.46 (s, 9H, C(CH₃)₃), 1.58 (t, J = 12.1 Hz, 1H, H-2a), 1.59 (t, J = 12.6 Hz, 1H, H-6a), 2.70 (m, 1H, H-6b), 2.80 (m, 1H, H-2b), 3.25 (q, J = 10.1 Hz, 1H, H-4), 3.53 (m, 1H, H-3), 3.71 (m, 1H, H-5), 3.79 (s, 3H, OMe), 3.98 (dd, J = 5.9, 12.7 Hz, 1H, H-1'a), 4.12 (dd, J = 5.5, 12.7 Hz, 1H, H-1'b), 4.42 (s, 2H, PhCH₂), 4.63 (m, 1H, NH), 5.27 (dd, J = 1.1, 10.4 Hz, 1H, H-3'Z), 5.18 (dd, J = 1.5, 17.2 Hz, 1H, H-3'E), 5.87 (ddt, J = 5.7, 10.5, 17.1 Hz, 1H, H-2'), 7.27–7.35 (m, 5H, C₆H₅); ¹³C NMR (125 MHz, CDCl₃): δ 28.3 (C(CH₃)₃), 38.2, 38.3 (C-2, C-6), 52.6 (OMe), 58.2 (C-5), 60.2 (C-4), 66.9 (CH₂Ph), 70.7 (C-1'), 73.7 (C-3), 77.9 (C-1), 79.7 (C(CH₃)₃), 117.4 (C-3'), 127.6, 127.9, 128.4, 137.3 (6C, C₆H₅), 134.6 (C-2'), 155.4 (NCO), 172.3 (CO); HRMS Calcd for C₂₃H₃₂N₄NaO₆ [M+Na]⁺: 483.2214. Found: 483.2217.

4.1.16. Methyl (1R,3R,4R,5S)-5-azido-1-benzyloxy-4-(tert-butoxycarbonylamino)-3-(2-hydroxyethoxy)-cyclohexane-1-carboxylate (22)

To a solution of **21** (488 mg, 1.06 mmol) in dioxane/water (3:1, 10 mL) were added 2,6-lutidine (247 μL, 2.12 mmol), OsO₄ (2.5% in *tert*-butanol, 275 μL, 0.053 mmol), and NaIO₄ (907 mg, 4.24 mmol) under argon. The reaction was stirred at rt for 4 h, then water

(5 mL) and DCM (20 mL) were added. The organic layer was separated, and the water phase was extracted by DCM (3 × 20 mL). The combined organic layers were washed with brine (20 mL) and dried over Na₂SO₄. The solvent was removed, and the remains were purified by MPLC on silica (1% gradient of MeOH in DCM) to afford the crude aldehyde intermediate (406 mg) as a yellow oil. The aldehyde was dissolved in MeOH (8 mL) under argon and sodium borohydride (66.4 mg, 1.76 mmol) was added in four portions at 0 °C. The solution was stirred for 45 min at 0 °C and quenched by the addition of saturated aqueous NH₄Cl (10 mL). The mixture was extracted with EtOAc (3 × 20 mL), and the combined organic layers were washed with brine (20 mL) and dried (Na₂SO₄). After concentration in vacuo, the residue was purified by MPLC on silica (1% gradient of MeOH in DCM) to give **22** (321 mg, 65%) as a colorless oil.

[α]_D –15.4 (c 0.80, CHCl₃); IR (film) 3365 (s, br, OH, NH), 2930 (m), 2098 (vs, N₃), 1695 (vs, CO), 1520 (m), 1455 (m), 1367 (s), 1301 (s), 1251 (m), 1204 (m), 1160 (s), 1095 (m) cm⁻¹; ¹H NMR (500 MHz, CDCl₃): δ 1.46 (s, 9H, C(CH₃)₃), 1.57–1.66 (m, 2H, H-2a, H-6a), 2.26 (s br, 1H, OH), 2.72 (ddd, J = 3.0, 4.1, 12.8 Hz, 1H, H-6b), 2.82 (ddd, J = 3.0, 4.1, 12.5 Hz, 1H, H-2b), 3.34 (q, J = 9.8 Hz, 1H, H-4), 3.50–3.54 (m, 2H, H-3, H-1'a), 3.60 (m, 1H, H-5), 3.70 (m, 2H, H-2'), 3.74 (m, 1H, H-1'b), 3.79 (s, 3H, OMe), 4.42 (s, 2H, PhCH₂), 4.65 (d, J = 6.8 Hz, 1H, NH), 7.27–7.36 (m, 5H, C₆H₅); ¹³C NMR (125 MHz, CDCl₃): δ 28.3 (C(CH₃)₃), 38.1 (C-2), 38.2 (C-6), 52.7 (OMe), 58.3 (C-5), 59.8 (C-4), 62.0 (C-2'), 67.0 (CH₂Ph), 71.0 (C-1'), 75.7 (C-3), 77.9 (C-1), 80.2 (C(CH₃)₃), 127.6, 127.9, 128.4, 137.2 (6C, C₆H₅), 155.8 (NCO), 172.4 (CO); ESI-MS Calcd for C₂₂H₃₂N₄NaO₇ [M+Na]⁺: 487.22. Found: 487.23.

4.1.17. Methyl (1R,3R,4R,5S)-5-acetamido-3-(2-acetoxyethoxy)-1-benzyloxy-4-(tert-butoxycarbonylamino)-cyclohexane-1-carboxylate (23)

Compound **22** (220 mg, 0.473 mmol) and Pd/C (10%, 25 mg) were suspended in dioxane/NEt₃ (10:1, 11 mL) and hydrogenated (1 atm H₂) at rt for 3 h. The suspension was filtered over Celite and concentrated. To a solution of the residue in pyridine (6 mL) was added acetic anhydride (3 mL) at 0 °C under argon and the mixture was stirred at rt for 16 h. The solution was co-evaporated with toluene (3 × 5 mL) and the residue was purified by MPLC on silica (1% gradient of MeOH in DCM) to yield **23** (158 mg, 64%) as a colorless oil.

¹H NMR (500 MHz, CDCl₃): δ 1.43 (s, 9H, C(CH₃)₃), 1.62–1.67 (m, 2H, H-2a, H-6a), 1.90, 2.08 (2 s, 6H, 2COCH₃), 2.67 (d, J = 12.5 Hz, 1H, H-6b), 2.81 (d, J = 13.9 Hz, 1H, H-2b), 3.51 (m, 1H, H-3), 3.54 (q, J = 9.8 Hz, 1H, H-4), 3.69 (ddd, J = 3.1, 6.6, 11.5 Hz, 1H, H-1'a), 3.79 (m, 1H, H-5), 3.81 (s, 3H, OMe), 3.82 (m, 1H, H-1'b), 4.17 (ddd, J = 3.2, 6.6, 12.0 Hz, 1H, H-2'a), 4.24 (ddd, J = 3.1, 5.9, 12.0 Hz, 1H, H-2'b), 4.36, 4.50 (A, B of AB, J = 10.8 Hz, 2H, PhCH₂), 4.75 (d, J = 7.0 Hz, 1H, 4-NH), 6.53 (d, J = 7.4 Hz, 1H, 5-NH), 7.27–7.35 (m, 5H, C₆H₅); ¹³C NMR (125 MHz, CDCl₃): δ 20.8, 23.1 (2COCH₃), 28.2 (C(CH₃)₃), 37.0 (C-2), 38.5 (C-6), 48.9 (C-5), 52.6 (OMe), 57.2 (C-4), 63.6 (C-2'), 66.8 (CH₂Ph), 67.3 (C-1'), 76.3 (C-3), 78.1 (C-1), 79.8 (C(CH₃)₃), 127.5, 127.7, 128.3, 137.5 (6C, C₆H₅), 157.3 (NCO), 170.0, 170.9, 172.4 (3CO); ESI-MS Calcd for C₂₆H₃₈N₂NaO₉ [M+Na]⁺: 545.25. Found: 545.24.

4.1.18. Methyl (1R,3R,4R,5S)-5-acetamido-1-benzyloxy-4-(tert-butoxycarbonylamino)-3-(2-hydroxyethoxy)-cyclohexane-1-carboxylate (24)

A solution of **23** (176 mg, 0.337 mmol) in MeOH (5 mL) was treated with 1 M methanolic NaOMe (0.5 mL) under argon at rt for 6 h. The reaction mixture was neutralized with acetic acid and concentrated. The residue was purified by MPLC on silica (1% gradient of MeOH in DCM) to give **24** (139 mg, 86%) as a colorless foam.

$[\alpha]_D -19.2$ (c 1.00, CHCl₃); IR (film) 3328 (s, br, OH, NH), 2930 (s), 1731 (s, CO), 1690 (vs, CO), 1662 (s, CO), 1537 (s), 1455 (m), 1367 (s), 1305 (s), 1250 (m), 1206 (m), 1162 (s), 1099 (s) cm⁻¹; ¹H NMR (500 MHz, CDCl₃): δ 1.42 (s, 9H, C(CH₃)₃), 1.64–1.69 (m, 2H, H-2a, H-6a), 1.91 (s, 3H, COCH₃), 2.45 (s, 1H, OH), 2.62 (d, $J = 12.5$ Hz, 1H, H-6b), 2.82 (d, $J = 11.5$ Hz, 1H, H-2b), 3.47 (dt, $J = 3.9, 10.2$ Hz, 1H, H-3), 3.51–3.57 (m, 2H, H-4, H-1'a), 3.67 (m, 2H, H-2'), 3.73 (m, 1H, H-1'b), 3.78 (s, 3H, OMe), 3.84 (m, 1H, H-5), 4.36, 4.47 (A, B of AB, $J = 10.8$ Hz, 2H, PhCH₂), 5.00 (d, $J = 8.0$ Hz, 1H, 4-NH), 6.51 (s br, 1H, 5-NH), 7.27–7.34 (m, 5H, C₆H₅); ¹³C NMR (125 MHz, CDCl₃): δ 23.1 (COCH₃), 28.3 (C(CH₃)₃), 37.4 (C-2), 38.7 (C-6), 48.6 (C-5), 52.7 (OMe), 58.0 (C-4), 62.1 (C-2'), 66.9 (CH₂Ph), 70.9 (C-1'), 76.9 (C-3), 78.1 (C-1), 80.0 (C(CH₃)₃), 127.6, 127.8, 128.4, 137.5 (6C, C₆H₅), 157.4 (NCO), 170.2, 172.5 (2CO); HRMS Calcd for C₂₄H₃₆N₂NaO₉ [M+Na]⁺: 503.2364. Found: 503.2358.

4.1.19. Methyl (1R,3R,4R,5S)-5-acetamido-1-benzyloxy-4-(tert-butoxycarbonylamino)-3-(2-tosyloxyethoxy)-cyclohexane-1-carboxylate (25)

To a solution of **24** (117 mg, 0.243 mmol) in DCM were subsequently added NEt₃ (10.1 μ L, 0.729 mmol), DMAP (10 mg) and *p*-toluenesulfonyl chloride (69.8 mg, 0.365 mmol) at 0 °C under argon. The mixture was stirred at rt for 4 h and then quenched by the addition of MeOH (0.3 mL). After evaporation of the solvents the residue was purified by MPLC on silica (1% gradient of MeOH in DCM) to yield **25** (130 mg, 84%) as a colorless oil.

$[\alpha]_D -14.2$ (c 1.04, CHCl₃); IR (film) 3326 (m, br, NH), 3066 (w), 2926 (s), 1731 (vs, CO), 1690 (vs, CO), 1532 (s), 1455 (m), 1366 (vs), 1304 (s), 1248 (m), 1206 (m), 1189 (s), 1177 (vs), 1097 (s) cm⁻¹; ¹H NMR (500 MHz, CDCl₃): δ 1.42 (s, 9H, C(CH₃)₃), 1.64–1.69 (m, 2H, H-2a, H-6a), 1.89 (s, 3H, COCH₃), 2.43 (s, 3H, PhCH₂), 2.61 (d, $J = 12.4$ Hz, 1H, H-6b), 2.72 (d, $J = 11.5$ Hz, 1H, H-2b), 3.44 (m, 1H, H-3), 3.49 (q, $J = 9.4$ Hz, 1H, H-4), 3.69 (m, 1H, H-1'a), 3.76 (m, 1H, H-1'b), 3.78–3.81 (m, 4H, H-5, OMe), 4.12 (m, 2H, H-2'), 4.33, 4.46 (A, B of AB, $J = 10.7$ Hz, 2H, PhCH₂), 4.70 (d, $J = 8.0$ Hz, 1H, 4-NH), 6.49 (d, $J = 7.7$ Hz, 1H, 5-NH), 7.27–7.34 (m, 7H, 2H of C₆H₄, C₆H₅), 7.78 (BB' of AA'BB', $J = 8.3$ Hz, 2H of C₆H₄); ¹³C NMR (125 MHz, CDCl₃): δ 21.6 (PhCH₂), 23.2 (COCH₃), 28.2 (C(CH₃)₃), 36.8 (C-2), 38.4 (C-6), 48.9 (C-5), 52.6 (OMe), 57.1 (C-4), 66.9, 67.0 (C-1', CH₂Ph), 69.2 (C-2'), 76.5 (C-3), 78.0 (C-1), 80.0 (C(CH₃)₃), 127.6, 127.8, 127.9, 128.4, 129.9, 132.9, 137.5, 144.9 (12C, C₆H₄, C₆H₅), 157.3 (NCO), 170.0, 172.6 (2CO); HRMS Calcd for C₃₁H₄₂N₂NaO₁₀S [M+Na]⁺: 657.2452. Found: 657.2450.

4.1.20. Methyl (1R,3R,4R,5S)-5-acetamido-3-(2-azidoethoxy)-1-benzyloxy-4-(tert-butoxycarbonylamino)-cyclohexane-1-carboxylate (26)

To a solution of **25** (129 mg, 0.204 mmol) in DMF (5 mL) were added 15-crown-5 (20.1 μ L, 0.102 mmol) and sodium azide (133 mg, 2.04 mmol). The resulting suspension was stirred for 22 h at 60 °C under argon, then diluted with EtOAc (20 mL) and filtered over Celite. After co-evaporation with toluene (2 \times 10 mL), the residue was purified by MPLC on silica (1% gradient of MeOH in DCM) to yield **26** (85.7 mg, 83%) as a colorless solid.

$[\alpha]_D -21.0$ (c 0.99, CHCl₃); IR (KBr) 3363 (m, br, NH), 2926 (m), 2103 (s, N₃), 1732 (vs, CO), 1688 (vs, CO), 1661 (vs, CO), 1532 (m), 1455 (m), 1366 (s), 1305 (vs), 1250 (m), 1206 (m), 1163 (s), 1096 (m) cm⁻¹; ¹H NMR (500 MHz, CDCl₃): δ 1.42 (s, 9H, C(CH₃)₃), 1.65 (t, $J = 12.4$ Hz, 1H, H-6a), 1.68 (t, $J = 12.1$ Hz, 1H, H-2a), 1.90 (s, 3H, COCH₃), 2.64 (d, $J = 12.7$ Hz, 1H, H-6b), 2.82 (d, $J = 12.2$ Hz, 1H, H-2b), 3.35 (m, 2H, H-2'), 3.49 (m, 1H, H-3), 3.57 (q, $J = 9.6$ Hz, 1H, H-4), 3.63 (m, 1H, H-1'a), 3.78 (m, 1H, H-1'b), 3.79 (s, 3H, OMe), 3.83 (m, 1H, H-5), 4.36, 4.48 (A, B of AB, $J = 10.8$ Hz, 2H, PhCH₂), 4.78 (d, $J = 8.1$ Hz, 1H, 4-NH), 6.58 (d, $J = 7.7$ Hz, 1H, 5-NH), 7.27–7.34 (m, 5H, C₆H₅); ¹³C NMR (125 MHz, CDCl₃): δ

23.1 (COCH₃), 28.2 (C(CH₃)₃), 37.2 (C-2), 38.4 (C-6), 49.0 (C-5), 51.0 (C-2'), 52.6 (OMe), 57.4 (C-4), 66.9 (CH₂Ph), 68.5 (C-1'), 76.7 (C-3), 78.1 (C-1), 79.8 (C(CH₃)₃), 127.6, 127.8, 128.4, 137.5 (6C, C₆H₅), 157.2 (NCO), 170.1, 172.6 (2CO); HRMS Calcd for C₂₄H₃₅N₅NaO₇ [M+Na]⁺: 528.2429. Found: 528.2430.

4.1.21. Methyl (1R,3R,4R,5S)-5-acetamido-1-benzyloxy-4-(tert-butoxycarbonylamino)-3-[2-(4-chlorobenzamido)ethoxy]-cyclohexane-1-carboxylate (27)

Compound **26** (29.0 mg, 57.5 μ mol) and Pd/C (10%, 10 mg) were suspended in MeOH (2.5 mL). NEt₃ (150 μ L) was added and the suspension was hydrogenated (1 atm H₂) at rt for 1 h. The mixture was filtered over Celite and concentrated. To a solution of the residue in DCM (2 mL) were subsequently added NEt₃ (31.8 μ L, 0.230 mmol), DMAP (5 mg), and *p*-chlorobenzoyl chloride (14.7 μ L, 0.115 mmol) at 0 °C under argon. The solution was stirred at rt for 2.5 h, then NEt₃ (50 μ L) and MeOH (0.5 mL) were added and stirring was continued for 15 min. The solvents were evaporated and the residue was purified by MPLC on silica (1% gradient of MeOH in DCM) to yield **27** (23.4 mg, 66%) as a colorless solid.

$[\alpha]_D -37.9$ (c 0.51, CHCl₃); IR (KBr) 3319 (m, br, NH), 2926 (m), 1732 (s, CO), 1690 (vs, CO), 1659 (vs, CO), 1537 (vs), 1486 (m), 1454 (m), 1366 (m), 1305 (s), 1250 (m), 1206 (m), 1161 (m), 1092 (s) cm⁻¹; ¹H NMR (500 MHz, CDCl₃): δ 1.33 (s, 9H, C(CH₃)₃), 1.66 (t, $J = 11.3$ Hz, 1H, H-2a), 1.68 (t, $J = 12.4$ Hz, 1H, H-6a), 1.93 (s, 3H, COCH₃), 2.62 (ddd, $J = 2.8, 3.4, 12.0$ Hz, 1H, H-6b), 2.80 (ddd, $J = 2.6, 3.9, 11.4$ Hz, 1H, H-2b), 3.43 (ddd, $J = 4.4, 9.9, 10.7$ Hz, 1H, H-3), 3.53 (m, 2H, H-2'), 3.56 (q, $J = 10.0$ Hz, 1H, H-4), 3.64 (m, 1H, H-1'a), 3.75–3.80 (m, 4H, H-1'b, OMe), 3.89 (m, 1H, H-5), 4.35, 4.43 (A, B of AB, $J = 10.7$ Hz, 2H, PhCH₂), 4.94 (d, $J = 8.9$ Hz, 1H, 4-NH), 6.41 (d, $J = 8.1$ Hz, 1H, 5-NH), 6.78 (t, $J = 5.2$ Hz, 1H, 2'-NH), 7.24–7.31 (m, 5H, C₆H₅), 7.37, 7.74 (AA', BB' of AA'BB', $J = 8.6$ Hz, 4H, C₆H₄); ¹³C NMR (125 MHz, CDCl₃): δ 23.2 (COCH₃), 28.2 (C(CH₃)₃), 37.6 (C-2), 38.7 (C-6), 40.1 (C-2'), 48.0 (C-5), 52.7 (OMe), 58.6 (C-4), 67.0 (CH₂Ph), 68.4 (C-1'), 76.6 (C-3), 78.2 (C-1), 79.8 (C(CH₃)₃), 127.7, 128.0, 128.4, 128.6, 128.6, 132.6, 137.1, 137.5 (12C, C₆H₄, C₆H₅), 157.2 (NCO), 166.3, 170.2, 172.3 (3CO); HRMS Calcd for C₃₁H₄₀ClN₃NaO₈ [M+Na]⁺: 640.2396. Found: 640.2400.

4.1.22. Methyl (1R,3R,4R,5S)-5-acetamido-1-benzyloxy-3-[2-(4-chlorobenzamido)ethoxy]-4-(2-fluoroacetamido)-cyclohexane-1-carboxylate (28)

A solution of **27** (30.5 mg, 49.3 μ mol) was treated with DCM/TFA (2:1, 0.75 mL) for 3 h at 0 °C under argon. After concentration in vacuo the residue was dissolved in DCM (3 mL). NEt₃ (68.4 μ L, 0.493 mmol), DMAP (5 mg), and 2-fluoroacetyl chloride (5.6 μ L, 74.0 μ mol) were subsequently added at 0 °C under argon and the solution was stirred for 2.5 h at rt. Then NEt₃ (20 μ L) and MeOH (0.5 mL) were added and stirring was continued for 10 min. After evaporation to dryness the residue was purified by MPLC on silica (1% gradient of MeOH in DCM) to give **28** (19.7 mg, 69%) as a yellowish solid.

$[\alpha]_D -38.3$ (c 0.84, CHCl₃); IR (KBr) 3292 (m, br, NH), 2927 (m), 1733 (s, CO), 1660 (vs, CO), 1652 (vs, CO), 1555 (vs), 1487 (m), 1455 (m), 1367 (m), 1310 (m), 1206 (m), 1150 (m), 1093 (s) cm⁻¹; ¹H NMR (500 MHz, CDCl₃): δ 1.67 (t, $J = 12.1$ Hz, 1H, H-2a), 1.72 (t, $J = 12.7$ Hz, 1H, H-6a), 1.93 (s, 3H, COCH₃), 2.62 (m, 1H, H-6b), 2.87 (m, 1H, H-2b), 3.44 (ddt, $J = 3.9, 7.6, 12.0$ Hz, 1H, H-2'a), 3.54–3.59 (m, 2H, H-3, H-1'a), 3.65 (m, 1H, H-2'b), 3.74–3.80 (m, 4H, H-1'b, OMe), 3.94 (q, $J = 9.8$ Hz, 1H, H-4), 4.04 (m, 1H, H-5), 4.37, 4.44 (A, B of AB, $J = 10.8$ Hz, 2H, PhCH₂), 4.47, 4.62 (A, B of ABX, $J = 14.4, 47.2$ Hz, 2H, FCH₂), 6.29 (d, $J = 8.5$ Hz, 1H, 5-NH), 6.77 (t, $J = 5.2$ Hz, 1H, 2'-NH), 6.94 (dd, $J = 1.9, 8.9$ Hz, 1H, 4-NH), 7.26–7.35 (m, 5H, C₆H₅), 7.38, 7.75 (AA', BB' of AA'BB', $J = 8.5$ Hz, 4H, C₆H₄); ¹³C NMR (125 MHz, CDCl₃): δ 23.0 (COCH₃), 37.5 (C-2),

38.8 (C-6), 40.0 (C-2'), 47.1 (C-5), 52.8 (OMe), 57.5 (C-4), 67.0 (CH₂Ph), 68.3 (C-1'), 76.5 (C-3), 78.1 (C-1), 79.8 (d, *J* = 186.9 Hz, CH₂F), 127.6, 128.0, 128.4, 128.6, 128.6, 132.5, 137.2, 137.7 (12C, C₆H₄, C₆H₅), 166.3, 169.4, 170.5, 172.2 (4CO); HRMS Calcd for C₂₈H₃₃ClFN₃NaO₇ [M+Na]⁺: 600.1883. Found: 600.1884.

4.1.23. Sodium (1*R*,3*R*,4*R*,5*S*)-5-acetamido-1-benzyloxy-3-[2-(4-chlorobenzamido)ethoxy]-4-(2-fluoroacetamido)-cyclohexane-1-carboxylate (**3**)

To a solution of **28** (16.5 mg, 28.5 μmol) in THF (4.5 mL) was added a solution of LiOH·H₂O (8 mg) in water (0.5 mL) at 0 °C. The mixture was stirred for 21 h at rt, then neutralized with acetic acid and concentrated in vacuo. The crude product was purified by reversed-phase chromatography (RP-18 column, 5% gradient MeOH in water), Dowex 50X8 ion-exchange chromatography (Na⁺ type), and P2 size exclusion chromatography to afford **3** (13.8 mg, 83%) as a colorless solid after a final lyophilization from water.

[α]_D²⁰ –19.9 (c 0.49, MeOH); IR (KBr) 3430 (vs, br, NH), 2927 (m), 1645 (vs, CO), 1598 (vs, CO), 1554 (s), 1487 (m), 1454 (m), 1385 (m), 1314 (m), 1093 (s), 1015 (m) cm⁻¹; ¹H NMR (500 MHz, D₂O): δ 1.44 (t, *J* = 11.8 Hz, 1H, H-2a), 1.58 (t, *J* = 12.5 Hz, 1H, H-6a), 1.87 (s, 3H, COCH₃), 2.42 (ddd, *J* = 2.8, 3.6, 12.4 Hz, 1H, H-6b), 2.79 (ddd, *J* = 2.8, 3.9, 11.9 Hz, 1H, H-2b), 3.50 (m, 2H, H-2'), 3.63 (dt, *J* = 4.4, 10.6 Hz, 1H, H-3), 3.69 (ddd, *J* = 4.4, 6.4, 10.6 Hz, 1H, H-1'a), 3.76–3.83 (m, 2H, H-4, H-1'b), 3.94 (ddd, *J* = 4.0, 11.1, 12.4 Hz, 1H, H-5), 4.30, 4.34 (A, B of AB, *J* = 10.3 Hz, 2H, PhCH₂), 4.46, 4.63 (A, B of ABX, *J* = 14.4, 46.4 Hz, 2H, FCH₂), 7.29–7.38 (m, 5H, C₆H₅), 7.47, 7.69 (AA', BB' of AA'BB', *J* = 8.6 Hz, 4H, C₆H₄); ¹³C NMR (125 MHz, D₂O): δ 21.8 (COCH₃), 37.9 (C-2), 38.5 (C-6), 40.0 (C-2'), 47.7 (C-5), 56.9 (C-4), 66.8 (CH₂Ph), 67.9 (C-1'), 76.4 (C-3), 79.3 (d, *J* = 182.7 Hz, CH₂F), 80.3 (C-1), 128.2, 128.6, 128.6, 128.7, 128.8, 131.9, 137.2, 137.7 (12C, C₆H₄, C₆H₅), 169.7, 171.0, 173.4, 177.2 (4CO); HRMS Calcd for C₂₇H₃₀ClFN₃Na₂O₇ [M+Na]⁺: 608.1546. Found: 608.1546.

4.2. In vitro binding assay

Murine MAG_{d1-3}-Fc was affinity purified from CHO-Lec 3.2.8.1 cell culture supernatant as described before⁴¹, dialyzed against 10 mM phosphate buffer pH 7.4, sterile filtered and stored at 4 °C. Inhibition assays for Siglecs were performed as described previously.^{26,32,41} In brief, fetuin was immobilized in microtiter plates and binding of MAG-Fc was determined in the presence of seven to eight different concentrations for each inhibitor using alkaline phosphatase-labeled anti-Fc antibodies. The half maximal inhibitory concentrations (IC₅₀) were determined from corresponding binding curves. In order to compare the results from different assays compound **29** was included in each test and used as a reference to calculate the relative inhibitory concentrations (rIC₅₀). At least four independent titrations were performed for each compound.

4.3. Surface plasmon resonance experiments (Biacore)³¹

Surface plasmon resonance experiments were performed on a Biacore 3000 machine using CM5 chips. Goat anti-human IgG antibody (Fc specific, Sigma) was immobilized following the standard Biacore EDC/NHS immobilization procedure. For this, 2 μL of a Fc-antibody solution of 2.1 mg/mL (in 50 mM phosphate buffer, pH 7.0) was added to 98 μL of 10 mM acetate buffer (pH 5.5). The immobilization yielded between 4.000 and 5.000 resonance units. A sample and a reference surface were prepared sequentially. For capturing MAG_{d1-3}-Fc onto the sample surface, MAG_{d1-3}-Fc in HBS-EP buffer (10 mM HEPES pH 7.4, 0.15 M NaCl, 3 mM EDTA, 0.0005% surfactant P20; Biacore) was injected. Ligands were dissolved in HBS-EP buffer. Eight ligand concentrations

of a twofold dilution series starting from 50 μM were prepared. The samples were injected in a randomized order. Five blank buffer injections were performed before triplicate measurements and one between each single experiment. The flow rate was 10 μL/min for the immobilization, 1 μL/min for the capturing of MAG, and 20 μL/min for the injection of ligands. For the determination of K_D values and for the kinetics analysis, standard Biacore software, version 3.2, was used.

Acknowledgment

We thank the Volkswagen Foundation, the Swiss National Science Foundation and the German Federal Ministry for Education and Research (BMBF, project 031632A), and the Tönjes-Vagt Foundation (project XXI) for their support of this work.

Supplementary data

Supplementary data (HPLC data and NMR spectra of target compound **3**) associated with this article can be found, in the online version, at doi:10.1016/j.bmc.2010.08.027.

References and notes

- Crocker, P. R.; Paulson, J. C.; Varki, A. *Nat. Rev. Immunol.* **2007**, *7*, 255.
- Kelm, S.; Pelz, A.; Schauer, R.; Filbin, M. T.; Tang, S.; de Bellard, M. E.; Schnaar, R. L.; Mahoney, J. A.; Hartnell, A.; Bradfield, P. *Curr. Biol.* **1994**, *4*, 965.
- Schwab, M. E.; Bandtlow, C. E. *Nature* **1994**, *371*, 658.
- Ramon y Cajal, S. *Degeneration and Regeneration of the Nervous System*; Oxford Univ. Press: London, 1928.
- Schwab, M. E.; Caroni, P. *J. Neurosci.* **1988**, *8*, 2381.
- Sandvig, A.; Berry, M.; Barrett, L. B.; Butt, A.; Logan, A. *Glia* **2004**, *46*, 225.
- Filbin, M. T. *Nat. Rev. Neurosci.* **2003**, *4*, 703.
- He, Z. G.; Koprivica, V. *Ann. Rev. Neurosci.* **2004**, *27*, 341.
- Caroni, P.; Savio, T.; Schwab, M. E. *Prog. Brain Res.* **1988**, *78*, 363.
- Quarles, R. H. *Neurochem. Res.* **2009**, *34*, 79.
- Lauren, J.; Hu, F.; Chin, J.; Liao, J.; Airaksinen, M. S.; Strittmatter, S. M. *J. Biol. Chem.* **2007**, *282*, 5715.
- Robak, L. A.; Venkatesh, K.; Lee, H.; Raiker, S. J.; Duan, Y.; Lee-Osbourne, J.; Hofer, T.; Mage, R. G.; Rader, C.; Giger, R. J. *J. Neurosci.* **2009**, *29*, 5766.
- Yang, L. J.; Zeller, C. B.; Shaper, N. L.; Kiso, M.; Hasegawa, A.; Shapiro, R. E.; Schnaar, R. L. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 814.
- Tang, S.; Shen, Y. J.; DeBellard, M. E.; Mukhopadhyay, G.; Salzer, J. L.; Crocker, P. R.; Filbin, M. T. *J. Cell Biol.* **1997**, *138*, 1355.
- Vinson, M.; Srijibos, P. J. L. M.; Rowles, A.; Facci, L.; Moore, S. E.; Simmons, D. L.; Walsh, F. S. *J. Biol. Chem.* **2001**, *276*, 20280.
- Wörter, V.; Schweigreiter, R.; Kinzel, B.; Mueller, M.; Barske, C.; Böck, G.; Frentzel, S.; Bandtlow, C. E. *PLoS One* **2009**, *4*, e5218. doi:10.1371/journal.pone.0005218.
- Yang, L. J. S.; Lorenzini, I.; Vajn, K.; Mountney, A.; Schramm, L. P.; Schnaar, R. L. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 11057.
- Vyas, A. A.; Blixt, O.; Paulson, J. C.; Schnaar, R. L. *J. Biol. Chem.* **2005**, *280*, 16305.
- Collins, B. E.; Kiso, M.; Hasegawa, A.; Tropak, M. B.; Roder, J. C.; Crocker, P. R.; Schnaar, R. L. *J. Biol. Chem.* **1997**, *272*, 16889.
- Collins, B. E.; Ito, H.; Sawada, N.; Ishida, H.; Kiso, M.; Schnaar, R. L. *J. Biol. Chem.* **1999**, *274*, 37637.
- Ito, H.; Ishida, H.; Collins, B.; Fromholt, S.; Schnaar, R. L.; Kiso, M. *Carbohydr. Res.* **2003**, *338*, 1621.
- Schwizer, D.; Gächje, H.; Kelm, S.; Porro, M.; Schwardt, O.; Ernst, B. *Bioorg. Med. Chem.* **2006**, *14*, 4944.
- Schwardt, O.; Gächje, H.; Vedani, A.; Mesch, S.; Gao, G.; Spreafico, M.; von Orelli, J.; Kelm, S.; Ernst, B. *J. Med. Chem.* **2009**, *52*, 989.
- Bhunia, A.; Schwardt, O.; Gächje, H.; Gao, G.; Kelm, S.; Benie, A. J.; Hricovini, M.; Peters, T.; Ernst, B. *ChemBioChem* **2008**, *9*, 2941.
- Shin, S.; Gächje, H.; Schwardt, O.; Gao, G.; Ernst, B.; Kelm, S.; Meyer, B. *ChemBioChem* **2008**, *9*, 2946.
- Kelm, S.; Brossmer, R.; Isecke, R.; Gross, H. J.; Strenge, K.; Schauer, R. *Eur. J. Biochem.* **1998**, *255*, 663.
- Kelm, S.; Brossmer, R. PCT Patent WO 03/000709A2, 2003.
- Strenge, K.; Schauer, R.; Bovin, N.; Hasegawa, A.; Kiso, M.; Kelm, S. *Eur. J. Biochem.* **1998**, *258*, 677.
- Gao, G.; Smiesko, M.; Schwardt, O.; Gächje, H.; Kelm, S.; Vedani, A.; Ernst, B. *Bioorg. Med. Chem.* **2007**, *15*, 7459.
- Mesch, S.; Lemme, K.; Koliwer-Brandl, H.; Strasser, D. S.; Schwardt, O.; Kelm, S.; Ernst, B. *Carbohydr. Res.* **2010**, *345*, 1348.
- Mesch, S.; Moser, D.; Strasser, D.; Kelm, A.; Cutting, B.; Rossato, G.; Vedani, A.; Koliwer-Brandl, H.; Wittwer, M.; Rabbani, S.; Schwardt, O.; Kelm, S.; Ernst, B. *J. Med. Chem.* **2010**, *53*, 1597.

32. Shelke, S. V.; Gao, G. P.; Mesch, S.; Gähje, H.; Kelm, S.; Schwardt, O.; Ernst, B. *Bioorg. Med. Chem.* **2007**, *15*, 4951.
33. (a) Bieberich, E.; Liour, S. S.; Yu, R. K. *Methods Enzymol.* **2000**, *312B*, 339; (b) Kopitz, J.; Sinz, K.; Brossmer, R.; Cantz, M. *Eur. J. Biochem.* **1997**, *248*, 527.
34. May, A. P.; Robinson, R. C.; Vinson, M.; Crocker, P. R.; Jones, E. Y. *Mol. Cell* **1998**, *1*, 719.
35. Crocker, R. P.; Klem, S. *The Siglec Family of I-type Lectins*; Wiley-VCH: Weinheim, 2000, Vol. IV, p 579.
36. Kim, C. U.; Lew, W.; Williams, M. A.; Liu, H.; Zhang, L.; Swaminathan, S.; Bischofberger, N.; Chen, M. S.; Mendel, D. B.; Tai, C. Y.; Laver, G.; Stevens, R. C. *J. Am. Chem. Soc.* **1997**, *119*, 681.
37. Rohloff, J. C.; Kent, K. M.; Postich, M. J.; Becker, M. W.; Chapman, H. H.; Kelly, D. E.; Lew, W.; Louie, M. S.; McGee, L. R.; Prisbe, E. J.; Schultze, L. M.; Yu, R. H.; Zhang, L. *J. Org. Chem.* **1998**, *63*, 4545.
38. Boullanger, P.; Maunier, V.; Lafont, D. *Carbohydr. Res.* **2000**, *324*, 97.
39. Yu, W.; Mei, Y.; Kang, Y.; Hua, Z.; Jin, Z. *Org. Lett.* **2004**, *6*, 3217.
40. Pozsgay, V.; Ekborg, G.; Sampathkumar, S.-G. *Carbohydr. Res.* **2006**, *341*, 1408.
41. Bock, N.; Kelm, S. In *Methods in Molecular Biology*; Brockhausen, I., Ed.; Humana Press Inc: Totowa, NJ, 2006; Vol. 347, p 359.
42. Shi, J. Y.; Blundell, T. L.; Mizuguchi, K. *J. Mol. Biol.* **2001**, *310*, 243.
43. (a) Deane, C. M.; Kaas, Q.; Blundell, T. L. *Bioinformatics* **2001**, *17*, 541; (b) Deane, C. M.; Blundell, T. L. *Protein Sci.* **2001**, *10*, 599.
44. Tripos Inc.: St. Louis, MO, 2007.
45. (a) Thompson, J. D.; Higgins, D. G.; Gibson, T. J. *Nucleic Acids Res.* **1994**, *22*, 4673; (b) Larkin, M. A.; Blackshields, G.; Brown, N. P.; Chenna, R.; McGettigan, P. A.; McWilliam, H.; Valentin, F.; Wallace, I. M.; Wilm, A.; Lopez, R.; Thompson, J. D.; Gibson, T. J.; Higgins, D. G. *Bioinformatics* **2007**, *23*, 2947; (c) <http://www.ebi.ac.uk/tools/clustalw2>.
46. General: Case, D. A.; Pearlman, D. A.; Caldwell, J. W. *AMBER 7*; University of California: San Francisco, 2002; AMBER force field: Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S.; Weiner, P. *J. Am. Chem. Soc.* **1984**, *106*, 765.
47. Mohamadi, F.; Richards, N. G. J.; Guida, W. C.; Liskamp, R.; Lipton, M.; Caufield, C.; Chang, G.; Hendrickson, T.; Still, W. C. *J. Comput. Chem.* **1990**, *11*, 440.
48. Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225.
49. Cramer, C.; Truhlar, D. J. *Comput. Aided Mol. Des.* **1992**, *6*, 629.
50. Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics* **1996**, *14*, 33.
51. *Desmond Molecular Dynamics System*, 2.0109; D. E. Shaw Research: New York, NY, 2009.
52. Lemieux, R. U. *Chem. Soc. Rev.* **1989**, *18*, 347.

Appendix 4: J. Med. Chem. 2010

Low Molecular Weight Antagonists of the Myelin-Associated Glycoprotein: Synthesis, Docking, and Biological Evaluation

Stefanie Mesch,[†] Delia Moser,[†] Daniel S. Strasser,[†] Antje Kelm,[‡] Brian Cutting,[†] Gianluca Rossato,[†] Angelo Vedani,[†] Hendrik Koliwer-Brandl,[‡] Matthias Wittwer,[†] Said Rabbani,[†] Oliver Schwardt,[†] Soerge Kelm,[‡] and Beat Ernst^{*†}

[†]*Institute of Molecular Pharmacy, Pharmazentrum, University of Basel, Klingelbergstrasse 50, CH-4056 Basel, Switzerland and* [‡]*Department of Physiological Biochemistry, University of Bremen, D-28334 Bremen, Germany*

Received August 10, 2009

The injured adult mammalian central nervous system is an inhibitory environment for axon regeneration due to specific inhibitors, among them the myelin-associated glycoprotein (MAG), a member of the Siglec family (sialic-acid binding immunoglobulin-like lectin). In earlier studies, we identified the lead structure **5**, which shows a 250-fold improved in vitro affinity for MAG compared to the tetrasaccharide binding epitope of GQ1b α (**1**), the best physiological MAG ligand described so far. By modifying the 2- and 5-position, the affinity of **5** could be further improved to the nanomolar range (\rightarrow **19a**). Docking studies to a homology model of MAG allowed the rationalization of the experimental binding properties. Finally, pharmacokinetic parameters (stability in the cerebrospinal fluid, logD and permeation through the BBB) indicate the drug-like properties of the high-affinity antagonist **19a**.

Introduction

The injured adult mammalian central nervous system (CNS) lacks the ability for axon regeneration,^{1,2} predominantly due to specific inhibitors expressed on residual myelin and on astrocytes recruited to the site of injury.^{3–7} Several inhibitor proteins have been identified, one of them being the myelin-associated glycoprotein (MAG).⁸ MAG is a transmembrane glycoprotein⁹ belonging to the family of the sialic acid-binding immunoglobulin like lectins, the so-called Siglecs.^{10,11} On the surface of neurons, MAG interacts with two classes of targets: proteins of the family of Nogo receptors (NgR^a)^{12,13} and the gangliosides GD1a and GT1b.^{11,14–16} Although the relative roles of gangliosides and NgRs as MAG ligands have yet to be resolved,^{8,17} in some systems, MAG inhibition is completely reversed by sialidase treatment, suggesting that MAG uses sialylated glycans as its major axonal ligands.¹⁸ Therefore, blocking MAG with potent antagonists may be a valuable therapeutic approach to enhance axon regeneration.

So far, the native carbohydrate ligand with the highest affinity to MAG is the ganglioside GQ1b α (Figure 1).¹⁹ Recently, MAG affinity of a partial structure of GQ1b α , the tetrasaccharide **1**, could clearly be correlated with its ability to reverse MAG-mediated inhibition of axonal outgrowth.²⁰

To reduce the structural complexity of tetrasaccharide **1** and, at the same time, improve its pharmacodynamic and pharmacokinetic properties, numerous MAG antagonists have been prepared. Because the affinity of a series of gangliosides indicated that not only the terminal Neu5Ac α (2–3)Gal structure is essential for MAG binding but also the internal sialic acids, the corresponding sialylated and sulfated analogues were synthesized.^{21,22} Furthermore, structural information obtained by trNOE NMR²³ and STD NMR²⁴ suggested that the Gal β (1–3)GalNAc core contributes to binding mainly by hydrophobic contacts. This was confirmed by the successful replacement of this disaccharide substructure by noncarbohydrate linkers.²⁵ In addition, the α (2–6)-linked sialic acid could also be replaced by lipophilic substituents.²² Finally, a pivotal simplification was reported by Kelm and Brossmer when they found that sialic acid derivatives modified in the 2- (e.g., **2**),^{26,27} 5- (e.g., **3**),^{26,27} or 9-position^{26,28} (e.g., **4**) exhibited enhanced antagonistic activity.²⁶ Combining the best modifications found for the 2²⁹- and 9-position³⁰ in one molecule led to antagonists, e.g., **5**, with affinities in the low micromolar range.³⁰

In this communication, we report a small library of MAG antagonists based on lead compound **5**. The binding properties, evaluated by a hapten binding assay, surface plasmon resonance (SPR), and competitive NMR experiments were rationalized by docking studies to a homology model of MAG. Finally, according to pharmacokinetic parameters, e.g., stability in the cerebrospinal fluid, the drug-likeness of the identified high-affinity antagonists could be demonstrated.

Results and Discussion

The sialic acid derivatives reported by Kelm and Brossmer²⁶ and our group^{29,30} exhibit MAG affinities in the low μ M range. An example is the neuraminic acid (Neu5Ac) derivative

*To whom correspondence should be addressed. Phone: 0041 267 15 51. Fax: 0041 267 15 52. E-mail: beat.ernst@unibas.ch.

^aBBB, blood–brain barrier; CHO, Chinese hamster ovary; ClAc, 2-chloroacetyl; DCE, 1,2-dichloroethane; DCM, dichloromethane; DMAP, 4-dimethylamino-pyridine; DMF, *N,N*-dimethylformamide; FAc, 2-fluoroacetyl; Gal, galactose; GalNAc, *N*-acetylgalactosamine; IgG, immunoglobulin G; K_D , dissociation constant; MS, molecular sieve; Neu5Ac, *N*-acetylneuraminic acid; NgR, Nogo receptor; NIS, *N*-iodosuccinimide; NMR, nuclear magnetic resonance; nosyl, 2-nitrobenzylsulfonyl; PAMPA, parallel artificial membrane permeation assay; PDB, Protein Data Bank; *i*PrOH, 2-propanol; RP, reversed phase; STD NMR, saturation transfer difference NMR; TfOH, trifluoroacetic acid; THF, tetrahydrofuran; TMS, trimethylsilyl; trNOE, transfer nuclear Overhauser enhancement; *p*-Ts, *p*-tolylsulfonyl.

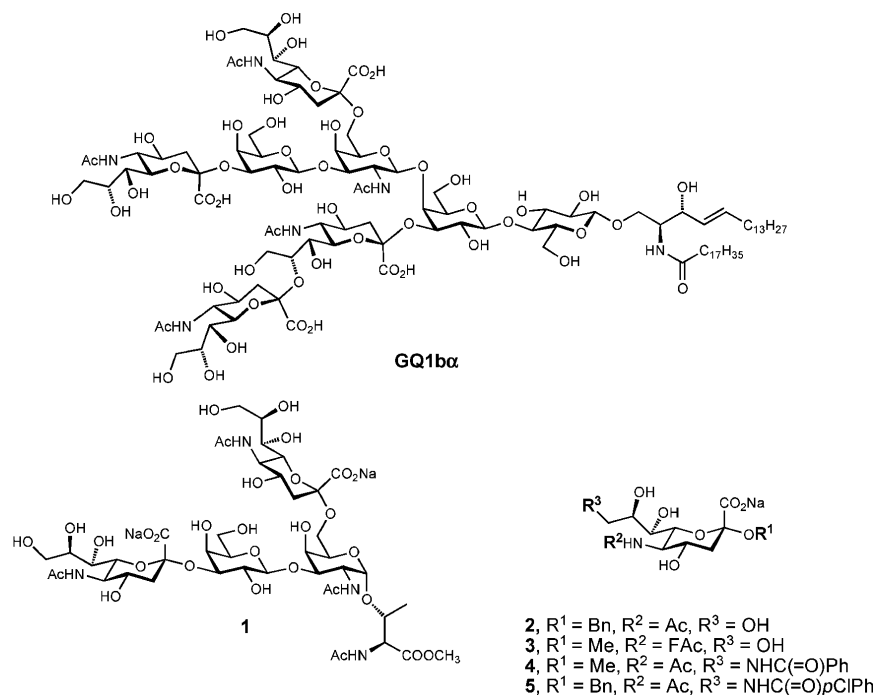


Figure 1. MAG antagonists; GQ1 α , the partial structure **1**,^{19,20} and sialic acid derivatives **2–5**.^{26,30}

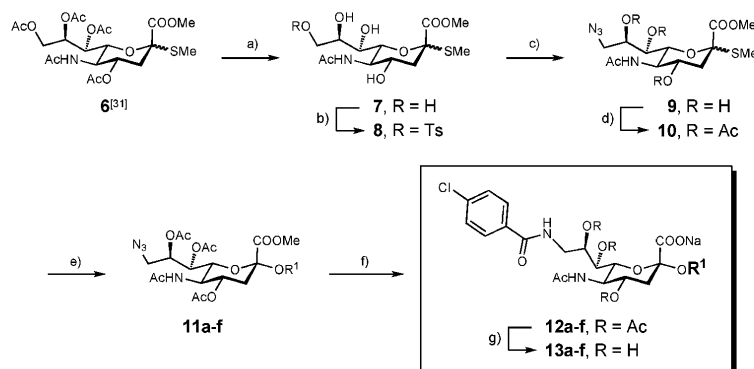
5³⁰ with a K_D of 17 μ M. Because a broad optimization effort for the 9-position had led to the identification of *p*-chlorobenzamide as the best substituent,³⁰ this position was not further investigated. The changeover from a methyl substituent in the 2-position to a benzyl group was rewarded by a 10-fold gain in affinity.²⁶ This effect can be rationalized by the results of STD-NMR investigations,²³ indicating a hydrophobic interaction of the α -face of D-galactose (see **1** in Figure 1) with MAG. When the hydrophobic contact was further extended by replacing the benzyl group with phenoxybenzyl or biphenyl substituents, only marginally improved affinities were detected.²⁹ Because the sheer enlargement of the hydrophobic group did not exhibit improved affinities, the electron density of the aglycone was altered in a next step. Finally, with the substituents in the 2- and 9-position set, a further optimization of the acyl group in the 5-position was conducted.^{26,27}

Synthesis of Sialosides 13a–f. Starting from Neu5Ac, thioglycoside **6** was synthesized according to a reported procedure.³¹ After deprotection by Zemplén conditions (\rightarrow **7**), the hydroxy group in the 9-position was selectively tosylated (\rightarrow **8**).³² Substitution of the tosylate using sodium azide and 15-crown-5 in DMF (\rightarrow **9**)³³ followed by acetylation yielded the sialyl donor **10**. For the introduction of the aglycone, **10** was then reacted with various benzyl alcohols (see Table 1, entries 2–9) in the presence of the promoters NIS/TfOH.³⁴ The sialosides **11a–f** were obtained as separable anomeric mixtures (α/β = 7/1 to 9/1). Amidation with *p*-chlorobenzoylchloride under modified Staudinger conditions³⁵ (\rightarrow **12a–f**) and final deprotection gave the amides **13a–f** in good yields (Scheme 1, Table 1).

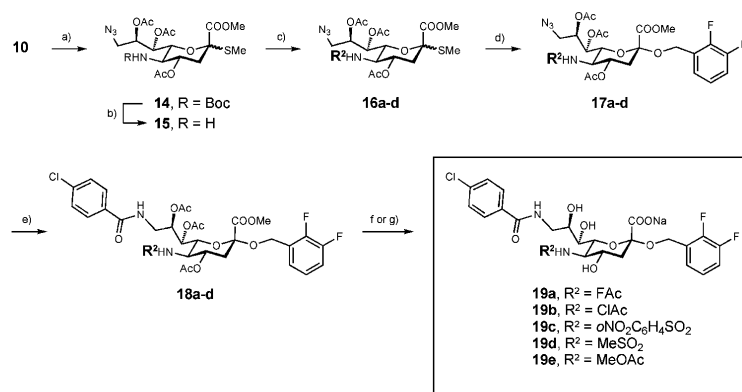
Synthesis of Sialosides 19a–g. As reported earlier, halogenation of the *N*-acetyl group at the 5-position increases the binding affinity toward MAG by a factor of 10–20.²⁶ Therefore, both the *N*-fluoroacetyl and *N*-chloroacetyl derivatives **19a** and **19b** were prepared. As sulfonamides adapt a different geometry³⁶ compared to the corresponding amides, **19c** and **19d** allow a further exploration of the binding site. Finally, the cyclopropyl and cyclobutyl derivatives **19f** and **19g** were synthesized in order to explore the possibility for extended hydrophobic contacts.

For the cleavage of the *N*-acetate group, **10** was treated with hydrazine in the presence of (Boc)₂O^{37,38} and subsequently reacylated (\rightarrow **14**). Deprotection with TMSCl and PhOH³⁹ (\rightarrow **15**), followed by acylation with carboxylic or sulfonic acid derivatives, yielded **16a–d**. Glycosylation using 2,3-difluorobenzyl alcohol (\rightarrow **17a–d**), amidation using modified Staudinger conditions³⁵ (\rightarrow **18a–d**) and final deprotection gave the test compounds **19a–d**. When methanolic NaOH was used in the final deprotection step of **18b**, a nucleophilic substitution occurred, leading to a 1:1 mixture of the desired **19b** and the methoxyacetamide derivative **19e** (Scheme 2).

For the synthesis of the two remaining test compounds **19f** and **19g**, an analogue approach starting from the thiosialoside **10** was accomplished. However, in contrast to the synthesis of **19a–e**, a different sequence of the modifications conducted at the 2-, 5-, and 9-position was performed (Scheme 3). Cleavage of the *N*-acetate (\rightarrow **14**) and amidation under modified Staudinger conditions yielded compound **20**. Next, *N*-deprotection followed by *N*-acylation (\rightarrow **22f** and **22g**) and benzylation (\rightarrow **23f** and **23g**) yielded, after final

Scheme 1^a

^a (a) NaOMe, MeOH (61%); (b) *p*-TsCl, pyridine (66%); (c) NaN₃, 15-crown-5, DMF (78%); (d) Ac₂O, DMAP, pyridine (73%); (e) R¹OH, NIS, TIOH, MeCN (**11a-f**, α-isomers: 54–76%; β-isomers: 8–11%); (f) PPh₃, *p*-chlorobenzoylchloride, DCE, rt (**12a-f**, 45–60%); (g) 10% aq NaOH; Dowex 50 × 8, Na⁺ form (**13a-f**, 39–70%).

Scheme 2^a

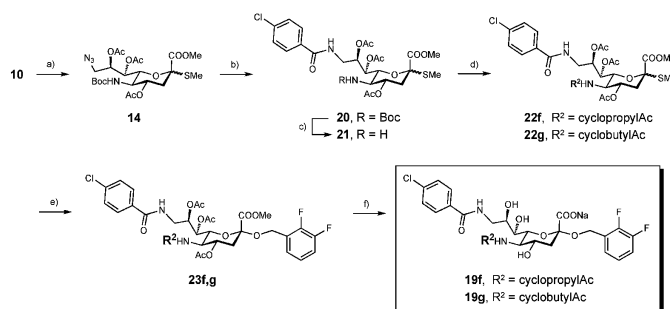
^a (i) Boc₂O, DMAP, THF, 60 °C, 4 h, (ii) N₂H₄ · H₂O, MeOH, rt, 24 h, (iii) Ac₂O, pyridine (76%); (b) 4 M PhOH, 4 M TMSCl in abs DCM (70%);³⁹ (c) acylation agent, NEt₃, DMAP, abs DCM, rt, 4 h or [ClCH₂C(=O)]₂O, NEt₃, dioxane/H₂O, rt (**16a-d**, 66–85%); (d) 2,3-difluorobenzyl alcohol, NIS, TIOH, MeCN (**17a-d**, α-isomers: 56–68%; β-isomers: 8–11%); (e) PPh₃, *p*-chlorobenzoylchloride, DCE, rt (**18a-d**, 48–58%); (f) 10% aq LiOH, THF/H₂O; Dowex 50 × 8, Na⁺ form (**19a-d**, 30–60%); (g) **18b**, 10% aq NaOH; Dowex 50 × 8, Na⁺ form (**19b**, 21%, **19e**, 19%).

deprotection, the desired cycloalkylacetic acid derivatives **19f** and **19g** in excellent overall yields.

Biological Evaluation. For the evaluation of the binding properties of the sialosides **13a-f** and **19a-g**, two previously reported assay formats were applied: a fluorescent hapten binding assay⁴⁰ and a SPR based biosensor (Biacore) experiment²² (Table 1). For the hapten inhibition assay, a recombinant protein consisting of the three *N*-terminal domains of MAG and the Fc part of human IgG (MAG_{d1-3}-Fc) was produced by expression in CHO cells and affinity purification on protein A-agarose.⁴⁰ The relative inhibitory concentrations (rIC₅₀) of the test compounds as competitive ligands were determined in microtiter plates coated with fetuin as the binding target for MAG_{d1-3}-Fc. By complexing the Fc-part with alkaline phosphatase-labeled anti-Fc antibodies and measuring the initial velocity of fluorescein release from fluorescein diphosphate, the amount of bound MAG_{d1-3}-Fc could be determined. At least three independent titrations

were performed for each compound tested with seven or eight concentrations in triplicates. The affinities were measured relative to the reference compound **5** (rIC₅₀ of 1, Table 1, entry 2). For the *K_D* determination in the Biacore assay, MAG_{d1-3}-Fc was immobilized on a dextran chip containing a surface of covalently bound protein A. A reference cell providing only protein A was used to compensate unspecific binding to the matrix.

By analyzing the affinities of **13a-f**, a substantial increase in affinity was achieved when the aromatic aglycone was halogenated in *ortho*- and *meta*-positions (entries 8 and 9). Less effective were halogens in the *para*-position (entries 4 and 5). In addition, with fluorine instead of chlorine, consistently slightly higher affinities (entries 4 and 8 vs entries 5 and 9) were obtained. With the symmetric pentafluorobenzyl group in **13c** (entry 6), an increase in affinity was expected, caused by an improved preorganization of the aglycone in the bioactive conformation. The observed

Scheme 3^a

^a(a) (i) Boc₂O, DMAP, THF, 60 °C, 4 h, (ii) N₂H₄·H₂O, MeOH, rt, 24 h, (iii) Ac₂O, pyridine (76%); (b) PPh₃, *p*-chlorobenzoylchloride, DCE, rt (48%); (c) 4 M PhOH, 4 M TMSCl in abs DCM (63%),³⁹ (d) R²COCl, NEt₃, DMAP, abs DCM, rt, 4 h (**22f**, 75%, **22g**, 39%); (e) 2,3-difluorobenzyl alcohol, NIS, TfOH, MeCN (**23f**, α: 73%, β: 8%; **23g**, α: 72%, β: 8%); (f) 10% aq LiOH, THF/H₂O: Dowex 50 × 8, Na⁺ form (**19f**, 44%; **19g**, 50%).

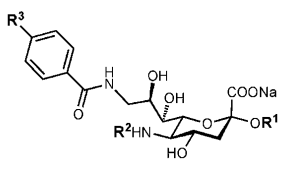
diminished affinity, i.e., a loss of a factor 3 compared to **13f**, may be the result of steric hindrance. Finally, when a 2-naphthylmethyl substituent was introduced (**13d**), a slight improvement of affinity compared to the benzyl substituent as present in **5** was obtained, presumably due to a favorable π - π interaction of the extended aromatic system. In a next step, substituent R² was optimized based on the so far most active antagonist **13f**. Kelm and Brossmer have demonstrated that halogenated acetamides in the 5-position of sialic acid derivatives strongly improve binding of MAG antagonists.²⁶ When this observation was applied to our series, the nanomolar fluoroacetamide **19a** (entry 10) was obtained. For chloroacetamide derivative **19b** (entry 11), the effect was less pronounced. Interestingly, an equally potent antagonist was achieved with the phenylsulfone substituent (**19c**, entry 12), while sulfone **19d** (entry 13) suffers from a drastic loss in activity. This is quite unexpected because an increase in the size of the acyl substituent was reported to lead to a reduction in affinity,²⁷ an observation that was confirmed by compounds **19e-g** (entries 14–16).























Stability in Cerebrospinal Fluid. For nerve regeneration, MAG antagonists will most likely be applied to the CNS by a local infusion. We therefore tested the stability of the fluoroacetate **19a** and the corresponding acetate **13f** in artificial cerebrospinal fluid (aCSF)⁴¹ and, as a control, in buffer solution for 19 h at 37 °C. According to LC-MS analysis, more than 95% of the initial concentrations of both antagonists were recovered from both media, predicting a high stability in the CNS, the target compartment of an *in vivo* application. Furthermore, logD_{octanol/water} values from -0.27 to 0.87 (see Table 1) might be beneficial for an intrathecal application because these distribution coefficients suggest a loss from the CNS compartment by a passive transport mechanism to be unlikely. This hypothesis is further supported by the results of the BBB-PAMPA assay showing log *P*_e values for **13f** and **19a** in the range of -10. For values below -5.7,⁴² no passive permeation through the BBB is expected.

Surface Plasmon Resonance (SPR). The interaction between MAG and MAG antagonists was analyzed by SPR experiments.^{43–45} For this purpose, MAG_{d1–3}-Fc was

immobilized on protein A, which on its part was covalently linked to a carboxymethyl dextrane surface of the chip. Whereas earlier Biacore investigations with MAG antagonists produced traditional sensorgrams,²² consistently negative sensorgrams, i.e., a net decrease in resonance units, were obtained with the compound series reported in Table 1. As an example, the sensorgram of **13f** is shown in Figure 2a. When fitted to a binding isotherm, these negative sensorgrams appear to clearly result from specific receptor–ligand interactions. To elucidate the origin of this unusual result, numerous factors such as the buffer capacity, the ion strength of the buffer, type, and matrix of the sensor chip as well as the applied type of immobilization of MAG_{d1–3}-Fc were analyzed.

As a result of ionizable functional groups of the analytes, pH variabilities could occur, a phenomenon previously reported by Mannen et al.⁴⁶ To avoid this effect, a sufficient buffer capacity, 50 mM instead of 10 mM HEPES, was applied. Furthermore, to exclude ionic repulsion effects, measurements were carried out at increased salt concentrations (150–500 mM NaCl). In addition, the effect of potential nonspecific binding to the dextran matrix or to protein A was analyzed by adding either dextran (1 mg/mL) to the buffer system or by conducting the experiment on a regenerated protein A surface. Because all these modifications of assay parameters did not provide an indication for the cause of the negative sensorgrams, different dextran biosensor chips, varying in carboxylate density (CM5 vs CM4), were analyzed as well. Although the reduction in signal intensity correlated well with the degree of functionalization of the chip surface, no influence on the sign of the sensorgrams could be detected (see Table S1 in Supporting Information). A further explanation for the negative sensorgrams could be a ligand-induced conformational change of the immobilized receptor leading to a decrease of its hydrodynamic radius and, as a consequence, yielding a negative refraction index.^{47,48} Because the negative refraction index correlates with the analyte concentration, we mirrored the negative sensorgrams, for an example see the sensorgram of **13f** in Figure 2b, to obtain SPR-determined equilibrium dissociation constants (*K*_{DS}, see Table 1).

Table 1. Relative Inhibitory Concentrations (rIC₅₀) Relative to Reference Compound **5**, K_D Values, and logD_{7.3} of Sialosides **13a–f** and **19a–g**


Entry	Compound	R ¹	R ²	R ³	rIC ₅₀ ^{a)}	K _D [μM]	logD (pH 7.3)
1	4 ^{26–28}	-CH ₃	Ac	H	18.00	137 ^{b)}	n.d. ^{c)}
2	5 ³⁰		Ac	Cl	1.00	17	n.d. ^{c)}
3	24 ³⁰		Ac	H	1.50	26	n.d. ^{c)}
4	13a		Ac	Cl	1.30	15	0.36
5	13b		Ac	Cl	1.20	13	-0.11
6	13c		Ac	Cl	0.26	6.1	-0.11
7	13d		Ac	Cl	0.74	11.6	0.58
8	13e		Ac	Cl	0.50	4.3	0.53
9	13f		Ac	Cl	0.30	2.4	-0.27
10	19a			Cl	0.02	0.5	-0.26
11	19b			Cl	0.07	2.1	0.35
12	19c			Cl	0.05	1.4	0.87
13	19d			Cl	0.60	17	-0.17
14	19e			Cl	0.14	2.3	0.06
15	19f			Cl	0.10	4.1	0.31
16	19g			Cl	0.14	5.8	0.75

^a rIC₅₀ is the concentration when 50% of the protein are inhibited, measured relative to reference compound **5**. ^b The affinity of compound **4** was measured using different protein batches, resulting in K_Ds of 137 μM and 105 μM (Table 2). All affinity data given in Table 1 were obtained with the protein batch showing a K_D of 137 μM for compound **4**; ^c n.d. not determined

To justify this procedure, we analyzed whether K_Ds obtained by mirroring the negative sensorgrams can be correlated with rIC₅₀ values determined by the fluorescent hapten

binding assay⁴⁰ (Figure 3). K_Ds from previously reported compounds^{25,30} (see Table S2, Supporting Information), which exhibit much higher molecular weights and therefore

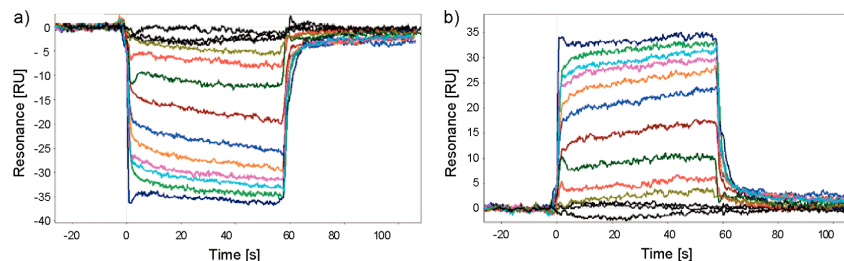


Figure 2. (a) Biacore sensorgrams for **13f** after subtraction of the reference; (b) mirrored sensorgrams of **13f**.

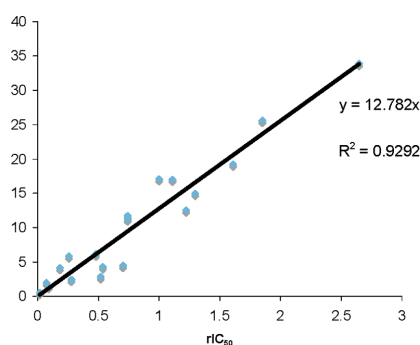


Figure 3. Correlation of K_D values obtained by SPR measurements with rIC_{50} values determined in the competitive inhibition assay,⁴⁰ Compounds **5**, **13a–f**, **19a–g**, and the compounds **25–29** from Table S2 (Supporting Information) were included.

generate positive sensorgrams, were also included in this correlation. The obtained correlation factor R^2 of 0.93 clearly suggests that the mirroring procedure does not falsify the binding information.

Determination of Relative Affinities by NMR. A further argument for the acceptance of the above-described mirroring of the sensorgrams was accomplished by competitive NMR experiments.⁴⁹ The approach is based on the molecular weight dependence of selective inversion recovery experiments (sTI). In the absence of binding, a selectively inverted NMR signal of a ligand requires a relatively long time to recover back to equilibrium. By contrast, if the ligand binds to a receptor, the time required to recover back to equilibrium is reduced. As a result, the binding of a ligand to a receptor is detectable through sTI experiments. Furthermore, these sTI experiments can be used for the ranking of the affinities of ligands relative to a reference compound.

For our purpose, the binding of antagonist **4** to MAG_{d1-3} -Fc was used as reference (K_D determined by Biacore is $137 \mu M$), whereas compounds **13f** and **25**²² (for the structure, see Supporting Information) were chosen as competitors because they fulfill two criteria. First, SPR experiments with both **13f** and **25** gave comparable K_D values, 2.4 and $2.8 \mu M$, respectively, and therefore require comparable concentrations for the observation of competitive binding. Second, the observed sensorgrams of compounds **13f** and **25** are of opposite sign, negative and positive, respectively.

In a first NMR experiment, it was demonstrated that **4** binds to MAG_{d1-3} -Fc according to the large differences

between selective inversion recovery of the *para*-hydrogen of the benzamide substituent in the presence or absence of MAG_{d1-3} -Fc (Figure 4a). For a quantitative evaluation of the relative affinities of **13f** and **25**, a titration curve describing the concentration dependence of the selective inversion recovery time of **4** was required (Figure 4b). The selective inversion time constants (sTI) were fit to a one-site binding model.

On the basis of the titration curve shown in Figure 4b, the determination of the relative affinities of **13f** and **25** became possible. With a NMR sample consisting of MAG_{d1-3} -Fc and compound **4** ($500 \mu M$), sTI for the initial point of the titration curve was remeasured. The obtained sTI of 1.03 ± 0.08 s compared to 1.05 ± 0.06 s for the first sample indicated a high degree of reproducibility. In a second step, $25 \mu M$ of either compound **13f** or **25** were added. Then, the sTIs were measured and the apparent concentration of compound **4** determined. When $25 \mu M$ of compound **13f** were added, the sTI increased to 1.84 ± 0.06 s, indicating an apparent concentration of compound **4** of 1.36 mM. Subtracting the actual concentration of compound **4** from its apparent concentration and dividing the result by the concentration of the inhibitor yields a relative affinity of 34.4 ± 2.1 for compound **13f** with respect to the reference compound **4** with the relative affinity of 1. With the identical procedure, a relative affinity of 17.6 ± 1.0 for compound **25** with respect to compound **4** was found (Figure 5).

The competitive NMR assay demonstrates that compounds **13f** and **25** both bind to MAG_{d1-3} -Fc with an affinity more than 1 order of magnitude greater than reference compound **4**. The relative affinities displayed in Figure 5 are in good agreement with the corresponding values determined by Biacore and the hapten inhibition assay.⁴⁰ Allowing a factor of 2 in the estimation of a compound's affinity with a single technology implies that the comparison of affinities between two assay formats may differ by as much as a factor of 4. Because for any of the three assays employed the affinities of **13f** and **25** to MAG_{d1-3} -Fc are within this range, the mirroring of the negative sensorgrams for the determination of K_D values is further justified.

Determination of Enthalpic and Entropic Contributions to Binding. For the elucidation of the thermodynamic parameters of the MAG /antagonist interaction, K_D s were measured in the Biacore assay at different temperatures. The analysis of a library of ligands containing structural modifications at the 2-, 5-, and 9-position of the neuraminic acid scaffold should allow the assignment of the enthalpic and entropic contributions to the various structural elements. The K_D values were determined at six different temperatures, starting at $5^\circ C$ and elevating the temperature by $5^\circ C$

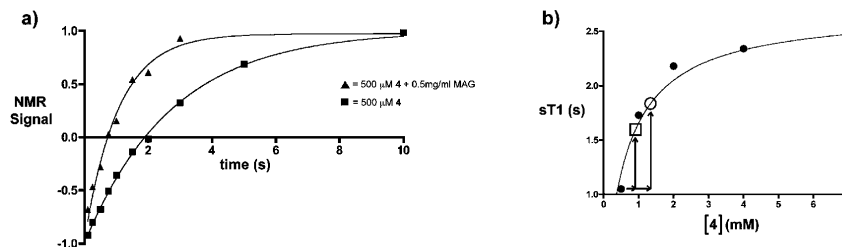


Figure 4. (a) Selective inversion recovery (sT1) of the *para*-hydrogen of the benzamide in compound **4**, either in the presence of MAG_{d1-3}-Fc (filled triangles) or in the absence of MAG_{d1-3}-Fc (filled squares). The normalized NMR signal represents the intensity of the *para*-hydrogen at a particular time divided by its intensity after 60 s of relaxation. (b) Titration of MAG_{d1-3}-Fc with compound **4**, and the observed sT1 (filled circles), observed sT1 when 500 μ M of **4** were mixed with 25 μ M **13f** (hollow circle), and observed sT1 when 500 μ M of **4** were mixed with 25 μ M **25** (hollow square). Vertical and horizontal arrows indicate the extent of attenuated relaxation of **4** when mixed with 25 μ M or competitor **13f** or **25** and the apparent concentration of **4**, respectively.

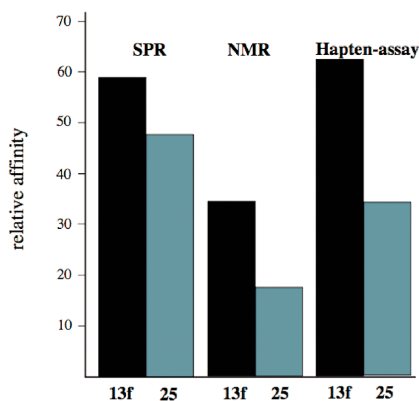


Figure 5. Affinity of compounds **13f** (black) and **25** (gray), relative to the affinity of compound **4** (= relative affinity of 1). The relative affinities are determined by NMR, Biacore, and the fluorescent hapten assay.⁴⁰

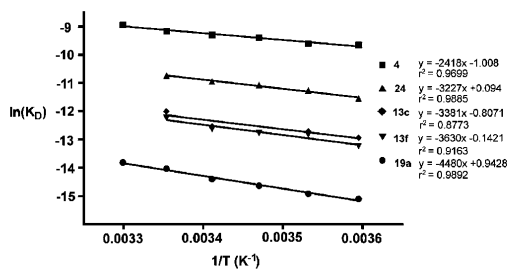


Figure 6. Van't Hoff plot. Measured data (dots) and corresponding linear fits according to eq 1.

$$\ln K_D = \frac{\Delta H}{RT} - \frac{\Delta S}{R} \quad (1)$$

steps up to 30 °C. The values were fitted according to eq 1 (Figure 6) leading to ΔH and ΔS^{50} (Table 2).

The analysis (Table 2) revealed that the improvement of the binding energies ΔG resulted mainly from enhanced binding enthalpies ΔH . The substitution of the methoxy group at

Table 2. Weighting of ΔH and ΔS with Respect to ΔG

entry	compd	ΔH [kJ/mol]	$-T^0\Delta S^*$ [kJ/mol]	ΔG [kJ/mol]	K_D [μ M]
17	4 ³⁰	-20.09	-2.52	-22.61	106 ^b
18	24 ³⁰	-26.85	0.25	-26.60	22
19	13c	-28.13	-1.98	-30.1	6.1
20	13f	-30.31	-0.20	-30.52	2.4
21	19a	-37.33	2.42	-34.91	0.5

^a $T = 298.13$ K. ^b The affinity of compound **4** was measured using different protein batches, resulting in K_D s of 137 μ M and 106 μ M. For the data given in Table 2, the protein batch showing a K_D of 106 μ M for compound **4** was used.

position 2 in **4** by a benzyloxy group (**24**, entry 18)³⁰ increased binding enthalpy by more than 6 kJ/mol and at the same time caused substantial entropy costs upon binding. With the pentafluorobenzyloxy substituent (**13c**, entry 19), both ΔH and ΔS were improved. Apparently, the interaction of the reducing end substituent can be optimized with an electron-poor aromatic ring. When the 2,3-difluorinated benzyloxy substituent (**13f**, entry 20) was introduced, ΔG could be further improved, mainly by a favorable enthalpy change. On the other hand, entropy costs increased as a result of the asymmetric substitution. The binding energy of the most active compound **19a** (entry 21, *N*-acetate replaced by *N*-fluoroacetate) is mainly based on a further enthalpic improvement. Unfortunately, in this case entropy costs of more than 2 kJ/mol have to be compensated, probably due to a specific conformational orientation requested for an optimal interaction of the FAc group.

Structure–Affinity Relationships. To interpret the binding affinities at the molecular level, we performed molecular docking studies. In the absence of a crystal structure, we used a homology model of the ligand-binding domain of MAG, which was recently applied successfully to a series of MAG antagonists (Figure 7).²² The compounds were manually docked using the salt bridge between the carboxylate of the sialosides with Arg118 and the hydrogen bond of C(9)-NH to the backbone nitrogen of Thr128 as anchor points. Finally, the protein–ligand complexes were fully minimized in aqueous solution.

For validation purposes, 12 compounds were docked and their binding strength quantified. Because these compounds bind at the protein surface, the contribution of solvation and entropy are difficult to estimate from thermodynamic docking studies. We therefore performed a series of molecular-dynamical simulations (4.0×10^{-9} s at 300 K) to elucidate the

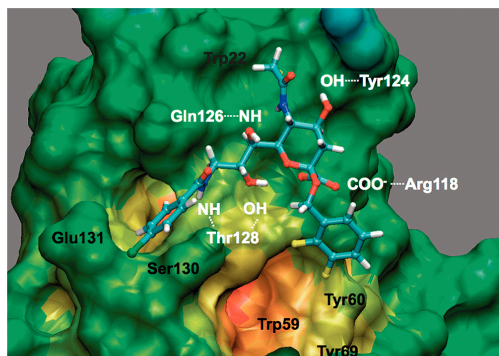


Figure 7. Homology model of MAG²² complexed with **19a** (most active compound of the series). Amino acids in white are forming hydrogen bonds and amino acids in black contribute to hydrophobic pockets. The salt bridge formed by the carboxylate group of **19a** with Arg118 and a hydrogen bond by C(9)-NH and the backbone nitrogen of Thr128 were used as anchor points for the docking. Hydrophobic interactions are established by the fluoroacetamido group and the side chains of Trp22 and Tyr124, the *p*-chlorobenzamide, and the side chains of Ser130 and Glu131, and the reducing end substituent, the 2,3-difluorobenzyl and the side chains of Trp59, Tyr60, and Tyr69, lining the main hydrophobic pocket. The image has been generated using VMD.⁵¹

kinetic aspects of binding and to quantify the contribution of hydrogen bonding over time. Details are given in the Supporting Information.

Upon analyzing the docking studies, the binding affinity could be associated to hydrophilic as well as hydrophobic interactions. The most important contribution is the salt bridge between the carboxylic acid and Arg118.^{26,52,53} Additionally, hydrogen-bond formations between 5-NH and the backbone carbonyl of Gln126, the carboxylate and the OH of Thr128, 8-OH and the backbone NH of Thr128 and 9-NH and the backbone carbonyl of Thr128 are observed. This latter finding is in good agreement with previous studies, where the abolishment of a hydrogen bond donor at position 9 resulted in a reduced binding affinity.²⁶ A considerable contribution to the binding affinity results from hydrophobic interactions. Thus, the *p*-chlorobenzamide is shown to point into a hydrophobic pocket, built by Ser130 and Glu131. A second hydrophobic pocket, which hosts the aglycone substituents, is defined by the side chains of Trp59, Tyr60, and Tyr69. With respect to different substitution patterns, the dichloro compound **13e** (Table 1, entry 8) shows a 4-fold and the difluoro compound **13f** (see Table 1, entry 9), a 7-fold enhancement in affinity compared to reference compound **5**, indicating a charge transfer complex with the electron rich aromatic ring of Tyr60. The only moderate improvement in binding affinity for the halogenated compounds **13a** and **13b** could be due to a steric clash of the *p*-substituent with the protein. Compound **13c** was synthesized as a symmetric analogon of compound **13f** in order to compensate entropy loss due to the orientation of the 2,3-difluorobenzyl ring. Again, that there is no improvement of the binding affinity in comparison to compound **13f** may be the consequence of a steric clash based on the *p*-substituent. Finally, the improved binding of **13d** might result from favorable π - π interactions of the naphthalene with Tyr69.

Some of the compounds modified at the 5-position seem to undergo a favorable σ - π interaction with Trp22. In case of **19a** and **19b**, we assume that the positively polarized hydrogens of the FAc or ClAc, respectively, stick favorably into the aromatic ring.⁵⁴ As fluorine is more electronegative, the polarization of the hydrogens is stronger and therefore the interaction is more favorable. For compounds **19f** and **19g**, additional hydrophobic interactions are possible, but the binding site seems to be spatially limited.²⁷ The reduced affinity of **19d** could be a consequence of the different bond angle for the sulfonamide substituent compared to an acetate in the same position (**13f**, entry 9), leading to different spatial requirements. Whereas methylsulfonamide **19d** shows a decrease in binding affinity, the nosyl substituent (**19c**) shows an opposite behavior. This might be due to the formation of a charge transfer complex with Trp22. To summarize, modifications at the reducing end improved binding affinity by a factor of 7 (**5**→**13f**). Combined with the best modification at the 5-position, the high affinity ligand **19a** was obtained.

Conclusion

In conclusion, the nanomolar affinity of the sialic acid derivative **19a** containing a difluorobenzyl substituent at the 2-, a fluoroacetate at the 5-, and *p*-chlorobenzamide at the 9-position clearly indicates the additivity of the beneficial effect of the various modifications. In addition, the thermodynamic analysis reveals that the improved affinity of **19a** predominantly results from an increased binding enthalpy and not from an entropy gain. The beneficial pharmacokinetic properties, e.g., a high stability in the cerebrospinal fluid, also support the drug-like properties of the newly identified MAG antagonist. However, due to the shallow binding site generally responsible for short half-lives of carbohydrate-protein complexes,^{22,55-58} it remains to be shown whether the complex formed by **19a** and MAG exhibits sufficient kinetic stability for in vivo applications.

Experimental Section

Chemistry. NMR spectra were recorded on a Bruker Avance DMX-500 (500 MHz) spectrometer. Assignment of ¹H and ¹³C NMR spectra was achieved using 2D methods (COSY, HSQC, TOCSY). Chemical shifts are expressed in ppm using residual CHCl₃, CHD₂OD, and HDO as references. Optical rotations were measured using Perkin-Elmer polarimeters 241 and 341. MS analyses were carried out using a Waters Micromass ZQ Detector system. The spectra were recorded in positive or negative ESI mode. The HPLC/HRMS analyses were carried out using a Agilent 1100 equipped with a photodiode array detector and a Micromass QTOF I equipped with a 4 GHz digital-time converter. All target compounds exhibit a purity of $\geq 95\%$. Reactions were monitored by TLC using glass plates coated with silica gel 60 F₂₅₄ (Merck) and visualized by using UV light and/or by charring with a molybdate solution (a 0.02 M solution of ammonium cerium sulfate dihydrate and ammonium molybdate tetrahydrate in aqueous 10% H₂SO₄). Column chromatography was performed on silica gel (Uetikon, 40-60 mesh). Methanol was dried by refluxing with sodium methoxide and distilled immediately before use. Pyridine was freshly distilled under argon over CaH₂. Dichloromethane (DCM), dichloroethane (DCE), acetonitrile (MeCN), toluene, and benzene were dried by filtration over Al₂O₃ (Fluka, type 5016 A basic). Molecular sieves (3 Å) were activated under vacuum at 500 °C for 2 h immediately before use. Compound **6** was prepared according to a published procedure.³¹ HPLC chromatograms and ¹H NMR spectra of the target compounds can be found in the Supporting Information.

Methyl (Methyl 5-Acetamido-3,5-dideoxy-2-thio-D-glycero- α -D-galacto-2-nonulopyranosid)onate (7). Compound **6** (217 mg, 42.0 mmol) was dissolved in dry MeOH (8.0 mL) and treated with NaOMe (1 M, 1.0 mL) for 2 h. The reaction mixture was neutralized with Amberlyst 15, filtered over a pad of celite, and the celite washed thoroughly with MeOH. The solvent was evaporated under reduced pressure, and the crude product was purified by chromatography on silica gel (1% gradient MeOH in DCM) to yield **7** as a white foam (90.0 mg, 61%). ¹H NMR (500 MHz, CD₃OD) δ 1.97 (dd, J = 11.5, 13.9 Hz, 1H, H-3a), 2.03 (s, 3H, SMe), 2.09 (s, 3H, NHAc), 2.47 (dd, J = 4.9, 13.9 Hz, 1H, H-3b), 3.54 (d, J = 9.4 Hz, 1H, H-7), 3.67 (dd, J = 5.6, 11.6 Hz, 1H, H-9a), 3.81–3.86 (m, 6H, H-5, H-8, H-9b, OMe), 4.11 (m, 2H, H-4, H-6). ¹³C NMR (CD₃OD) δ 11.2 (SMe), 22.7 (NHAc), 41.2 (C-3), 53.1 (OMe), 54.1 (C-5), 65.2 (C-9), 68.3 (C-4), 70.2 (C-7), 71.2 (C-8), 72.6 (C-6), 84.6 (C-2), 170.8, 175.0 (2 CO). ESI-MS calcd for C₁₃H₂₃NO₈S [M + Na]⁺ 376.10; found m/z 376.10.

Methyl (Methyl 5-Acetamido-3,5-dideoxy-2-thio-9-tosyl-D-glycero- α -D-galacto-2-nonulopyranosid)onate (8). To a solution of **7** (3.20 g, 9.07 mmol) in freshly distilled pyridine (80 mL) *p*-toluenesulfonyl chloride (1.90 g, 10.0 mmol) was added at 0 °C and the mixture was stirred for 2 h at 0 °C. Afterward tosyl chloride (0.70 g, 3.68 mmol) was added and the solution stirred continuously for 16 h at 5 °C. The reaction mixture was warmed to rt, methanol (20 mL) was added and stirring continued for 30 min. After evaporation of the solvents, the crude product was purified by chromatography on silica gel (DCM:MeOH, 19:1) to give **8** as a foam (3.00 g, 66%). ¹H NMR (500 MHz, CDCl₃) δ 1.79 (dd, J = 11.2, 13.4 Hz, 1H, H-3a), 1.96 (s, 3H, SMe), 1.99 (s, 3H, NHAc), 2.37 (s, 3H, CH₃), 2.74 (dd, J = 3.7, 13.2 Hz, 1H, H-3b), 3.25 (d, J = 9.5 Hz, 1H, H-6), 3.41 (d, J = 9.7 Hz, 1H, H-7), 3.66 (s, 3H, OMe), 3.98 (m, 1H, H-8), 4.10 (m, 1H, H-9a), 4.24 (m, 1H, H-9b), 6.94 (d, J = 7.2 Hz, 1H, NHAc), 7.28, 7.72 (AA', BB' of AA'BB', J = 8.2 Hz, 4H, CH_{ar}). ¹³C NMR (CDCl₃) δ 11.9 (SMe), 21.7 (CH₃), 23.0 (NHAc), 40.4 (C-3), 52.7 (C-5), 53.7 (OMe), 67.8, 68.6, 69.1 (C-4, C-7, C-8), 72.0 (C-6), 82.4 (C-2), 128.0, 130.0, 132.4, 145.1 (6 C-Ar), 170.4, 174.2 (2 CO). ESI-MS calcd for C₂₀H₂₉NO₁₀S₂ [M + Na]⁺ 530.11; found m/z 530.19.

Methyl (Methyl 5-Acetamido-9-azido-3,5,9-trideoxy-2-thio-D-glycero- α -D-galacto-2-nonulopyranosid)onate (9). Compound **8** (160 mg, 0.32 mmol) was dissolved in dry DMF (5 mL). NaN₃ (103 mg, 1.58 mmol) and 15-crown-5 (28.6 mg, 0.13 mmol) were added successively and the reaction mixture was stirred at 60 °C for 24 h. After filtration through a pad of celite, the solvent was evaporated and the residue was purified by chromatography on silica gel (gradient 1% MeOH in DCM) to yield **9** (96.0 mg, 78%). ¹H NMR (500 MHz, CD₃OD) δ 1.79 (m, 1H, H-3a), 2.03 (s, 3H, SMe), 2.17 (s, 3H, NHAc), 2.77 (dd, J = 4.5, 12.8 Hz, 1H, H-3b), 3.39 (dd, J = 6.2, 12.7 Hz, 1H, H-9a), 3.48 (m, 2H, H-6, H-7), 3.57 (dd, J = 2.3, 12.7 Hz, 1H, H-9b), 3.69 (m, 1H, H-4), 3.79 (m, 1H, H-5), 3.86 (s, 3H, OMe), 3.93 (ddd, J = 2.6, 5.7, 8.5 Hz, 1H, H-8), 4.01 (m, 1H, H-6). ¹³C NMR (CD₃OD) δ 12.0 (SMe), 22.7 (NHAc), 41.8 (C-3), 53.6 (C-5), 53.8 (OMe), 55.2 (C-9), 69.0 (C-4), 71.0 (C-7), 71.7 (C-8), 76.8 (C-6), 171.7, 175.2 (2 CO). ESI-MS calcd for C₁₃H₂₂N₄O₈S [M + Na]⁺ 401.11; found m/z 401.15.

Methyl (Methyl 5-Acetamido-4,7,8-tri-O-acetyl-9-azido-3,5,9-trideoxy-2-thio-D-glycero- α -D-galacto-2-nonulopyranosid)onate (10). Compound **9** (87.0 mg, 0.23 mmol) was dissolved in dry pyridine (1.0 mL) and cooled to 0 °C. After 15 min, DMAP (4.47 mg, 0.04 mmol) and Ac₂O (0.5 mL) were added successively. The reaction mixture was warmed to rt and stirred for 14 h. The solvent was evaporated and the residue purified by chromatography on silica gel (toluene/ethylacetate, 1:3) to afford **10** (85.0 mg, 73%). ¹H NMR (500 MHz, CDCl₃) δ 1.81 (s, 3H, NHAc), 1.91 (m, 1H, H-3a), 1.97, 2.05, 2.10 (3 s, 9H, 3 OAc), 2.13 (s, 3H, SMe), 2.67 (dd, J = 4.4, 12.5 Hz, 1H, H-3b), 3.19 (dd, J = 5.7, 13.4 Hz, 1H, H-9a), 3.59 (m, 1H, H-9b), 3.76 (m, 4H, H-6, OMe), 4.01 (m, 1H, H-5), 4.81 (m, 1H, H-4), 4.91 (d, J = 7.7 Hz,

1H, NHAc), 5.24 (m, 2H, H-7, H-8). ¹³C NMR (CDCl₃) δ 12.1 (SMe), 20.9, 21.1 (3C, 3 OAc), 23.2 (NHAc), 37.8 (C-3), 49.4 (C-5), 50.6 (C-9), 53.0 (OMe), 68.1 (C-4), 69.7 (C-7), 72.3 (C-6), 74.6 (C-8), 88.4 (C-2), 170.2, 170.3, 171.0 (5C, CO). ESI-MS calcd for C₁₉H₂₈N₄O₁₀S [M + Na]⁺ 527.40; found m/z 527.21.

General Procedure for the Synthesis of Compounds 11a–f, 17a–d, 18f.g. Compound **10** (0.12 mmol) was dissolved in dry acetonitrile (2.0 mL) under argon. The alcohol (0.26 mmol) and powdered MS 3 Å (80.0 mg) were added. The mixture was stirred at rt for 1.5 h. Then the suspension was cooled to –40 °C and subsequently treated with *N*-iodosuccinimide (0.60 mmol) and triflic acid (0.06 mmol in 0.2 mL MeCN). After 30 min, the reaction mixture was warmed to –30 °C and stirring continued for 16 h. The mixture was then warmed to rt, stirred for another 2 h, and filtered through a pad of celite. The celite was washed with DCM (10 mL), and the filtrate was subsequently washed with 20% aqueous Na₂S₂O₃ (1 mL) and saturated aqueous NaHCO₃ (3 \times 5 mL). The organic phase was dried over Na₂SO₄, filtered, and concentrated under reduced pressure. The crude product was purified by chromatography on silica gel.

General Procedure for the Synthesis of Compounds 12a–f, 20, 18a–d. Compounds **11a–f** (**14** or **17a–d**) (0.09 mmol) and *p*-chlorobenzoyl chloride (0.36 mmol) were dissolved in dry DCE (3.0 mL) under argon. Triphenylphosphine (0.18 mmol) in dry DCE (1.5 mL) was added after 5 min, and the solution was stirred at rt for 24 h. The reaction mixture was diluted with DCM (10 mL) and washed with saturated aqueous NaHCO₃ (3 \times 10 mL). The organic phase was dried over Na₂SO₄, filtered, and concentrated under reduced pressure. The residue was purified by chromatography on silica gel.

General Procedure for the Deprotection of 12a–f, 18b. Compound **12a–f** (**18b**) (0.06 mmol) was dissolved in MeOH (1.5 mL) and treated with 10% aqueous NaOH (0.3 mL). The reaction mixture was stirred at rt for 3 h. Then the reaction mixture was neutralized with 7% aqueous HCl (0.2 mL). The solvent was evaporated, and the crude product was purified by chromatography on RP-18.

General Procedure of the Deprotection of 18a–d,f–g. Compound **18a–d,f–g** (0.03 mmol) was dissolved in THF/H₂O (2 mL/0.5 mL) and was reacted with LiOH (0.28 mmol). The crude product was purified on RP-18 (10% gradient MeOH in water) followed by ion exchange chromatography (Dowex-50) and P2 size exclusion chromatography to yield **19**.

Methyl (4-Chlorobenzyl 5-Acetamido-4,7,8-tri-O-acetyl-9-azido-3,5,9-trideoxy-D-glycero- α -D-galacto-2-nonulopyranosid)onate (11a). Compound **10** (61.0 mg, 0.12 mmol) was reacted with 4-chlorobenzyl alcohol (38.0 mg, 0.26 mmol), *N*-iodosuccinimide (134 mg, 0.601 mmol), and triflic acid (6.00 μ L, 9.00 mg, 0.06 mmol). The crude product was purified by chromatography on silica gel (0.5% gradient *i*PrOH in petrol ether/DCM 8:4) to yield **11a** (55.0 mg, 76%) as a colorless oil. [α]_D²⁰ –0.02 (*c* 0.28, CH₂Cl₂). ¹H NMR (500 MHz, CDCl₃) δ 1.90 (s, 3H, NHAc), 2.03 (m, 4H, OAc, H-3a), 2.17, 2.20 (2 s, 6H, 2 OAc), 2.65 (dd, J = 4.6, 12.9 Hz, 1H, H-3b), 3.26 (dd, J = 5.7, 13.4 Hz, 1H, H-9a), 3.55 (dd, J = 2.8, 13.4 Hz, 1H, H-9b), 3.71 (s, 3H, OMe), 4.12 (m, 2H, H-5, H-6), 4.41, 4.76 (A, B of AB, J = 12.3 Hz, 2H, CH₂Ar), 4.87 (m, 1H, H-4), 5.15 (d, J = 9.8 Hz, 1H, NH), 5.35 (m, 2H, H-7, H-8) 7.30 (m, 4H, CH_{ar}). ¹³C NMR (CDCl₃) δ 20.8, 20.9, 21.1 (3 OAc), 23.3 (NHAc), 38.0 (C-3), 49.4 (C-5), 51.0 (C-9), 52.8 (OMe), 66.1 (CH₂Ar), 67.9 (C-7), 68.9 (C-4), 69.4 (C-8), 72.8 (C-6), 98.6 (C-2), 128.4, 129.1, 133.5, 135.7 (6 C-Ar), 168.3, 170.3, 170.3, 171.0 (5C, 5 CO). ESI-MS calcd for C₂₅H₃₁ClN₄O₁₁ [M + Na]⁺ 621.17; found m/z 621.24.

Methyl (4-Fluorobenzyl 5-Acetamido-4,7,8-tri-O-acetyl-9-azido-3,5,9-trideoxy-D-glycero- α -D-galacto-2-nonulopyranosid)onate (11b). Compound **10** (62.0 mg, 0.12 mmol) was reacted with 4-fluorobenzyl alcohol (38.0 mg, 0.26 mmol), *N*-iodosuccinimide (134 mg, 0.601 mmol), and triflic acid (6.00 μ L, 9.00 mg, 0.06 mmol). The crude product was purified by chromatography on silica gel (0.5% gradient *i*PrOH in petrol ether/DCM 8:4) to yield **11b** (53.0 mg, 76%) as a colorless oil. [α]_D²⁰

−0.01, (*c* 2.6, CH₂Cl₂). ¹H NMR (500 MHz, CDCl₃) δ 1.89 (s, 3H, NHAc), 2.02 (m, 1H, H-3a), 2.03, 2.17, 2.20 (3 s, 9H, 3 OAc), 2.65 (dd, *J* = 4.5, 12.8 Hz, 1H, H-3b), 3.29 (dd, *J* = 5.4, 13.3 Hz, 1H, H-9a), 3.58 (dd, *J* = 2.0, 13.1 Hz, 1H, H-9b), 3.71 (s, 3H, OMe), 4.13 (m, 2H, H-5, H-6), 4.41, 4.76 (A, B of AB, *J* = 11.9 Hz, 2H, CH₂Ar), 4.85 (m, 1H, H-4), 5.35 (m, 2H, H-8, H-7), 7.02 (t, *J* = 8.6 Hz, 2H, CH_{ar}), 7.31 (dd, *J* = 5.6, 8.2 Hz, 2H, CH_{ar}). ¹³C NMR (CDCl₃) δ 20.8, 20.9, 21.1 (3 OAc), 23.2 (NHAc), 49.3 (C-3), 51.0 (C-5), 52.8 (C-9), 53.5 (OMe), 66.1 (CH₂Ar), 68.0 (C-7), 68.9 (C-4), 69.7 (C-8), 72.8 (C-6), 98.5 (C-2), 115.2 (*J* = 21.7 Hz), 129.6 (*J* = 8.4 Hz), 132.9 (*J* = 2.9 Hz), 162.4 (*J* = 246.0 Hz) (6 C–Ar), 168.4, 170.2, 170.3, 170.4, 171.0 (6C, 6 CO). ESI-MS calcd for C₂₅H₃₁FN₄O₁₁ [M + Na]⁺ 605.20; found *m/z* 605.22.

Methyl (Pentafluorobenzyl 5-Acetamido-4,7,8-tri-O-acetyl-9-azido-3,5,9-trideoxy- α -D-galacto-2-nonulopyranosid)onate (11c). Compound **10** (60.0 mg, 0.12 mmol) was reacted with pentafluorobenzyl alcohol (60.0 mg, 0.30 mmol), *N*-iodosuccinimide (32.0 mg, 0.14 mmol), and triflic acid (4.00 μ L, 7.00 mg, 0.04 mmol). The crude product was purified by chromatography on silica gel (0.5% gradient *i*PrOH in petrol ether/DCM 8:4) to yield **11c** (47.0 mg, 66%) as a colorless oil. [α]_D²⁰ −0.03 (*c* 0.5, CH₂Cl₂). ¹H NMR (500 MHz, CDCl₃) δ 1.90 (s, 3H, NHAc), 1.94 (dd, *J* = 10.4, 12.6 Hz, 1H, H-3a), 2.03, 2.19, 2.21 (3 s, 9H, 3 OAc), 2.59 (dd, *J* = 4.6, 12.8 Hz, 1H, H-3b), 3.30 (dd, *J* = 5.6, 13.5 Hz, 1H, H-9a), 3.59 (dd, *J* = 2.9, 13.5 Hz, 1H, H-9b), 3.86 (s, 3H, OMe), 4.12 (m, 2H, H-5, H-6), 4.43 (A of AB, *J* = 11.0 Hz, 1H, CH₂Ar), 4.87 (m, 1H, H-4), 4.90 (B of AB, *J* = 10.7 Hz, 1H, CH₂Ar), 5.19 (d, *J* = 9.3 Hz, 1H, NH), 5.39 (m, 2H, H-7, H-8). ¹³C NMR (CDCl₃) δ 20.8, 20.9, 21.1 (3 OAc), 23.2 (NHAc), 37.7 (C-3), 49.4 (C-5), 51.0 (C-9), 53.0 (OMe), 54.0 (CH₂Ar), 67.8, 68.8, 69.3 (C-4, C-7, C-8), 72.9 (C-6), 98.7 (C-2), 167.8, 170.3 (5C, 5 CO). ESI-MS calcd for C₂₅H₂₇F₅N₄O₁₁ [M + Na]⁺ 677.16; found *m/z* 677.32.

Methyl (2-Naphthyl 5-Acetamido-4,7,8-tri-O-acetyl-9-azido-3,5,9-trideoxy- α -D-galacto-2-nonulopyranosid)onate (11d). Compound **10** (50.0 mg, 0.10 mmol) was reacted with 2-naphthalenemethanol (24.0 mg, 0.15 mmol), *N*-iodosuccinimide (27.0 mg, 0.12 mmol), and triflic acid (4.00 μ L, 7.00 mg, 0.04 mmol). The crude product was purified by chromatography on silica gel (0.5% gradient *i*PrOH in petrol ether/DCM 8:4) to yield **11d** (37.0 mg, 61%) as a colorless oil. ¹H NMR (500 MHz, CDCl₃) δ 1.85 (s, 3H, NHAc), 2.09 (m, 1H, H-3a), 2.03, 2.15, 2.21 (3 s, 9H, 3 OAc), 2.26 (dd, *J* = 6.0, 13.5 Hz, 1H, H-9a), 3.59 (dd, *J* = 2.9, 13.5 Hz, 1H, H-9b), 3.67 (s, 3H, OMe), 4.14 (m, 2H, H-5, H-6), 4.63 (A of AB, *J* = 12.2 Hz, 1H, CH₂Ar), 4.89 (m, 1H, H-4), 4.96 (B of AB, *J* = 12.2 Hz, 1H, CH₂Ar), 5.34 (m, 2H, H-7, NH), 5.38 (m, 1H, H-8), 7.45–7.48 (m, 3H, CH_{ar}), 7.79–7.83 (m, 4H, CH_{ar}). ¹³C NMR (CDCl₃) δ 21.3, 21.5 (3C, 3 OAc), 23.6 (NHAc), 38.5 (C-3), 49.9 (C-5), 51.3 (C-9), 53.1 (OMe), 67.4 (CH₂Ar), 68.4 (C-7), 69.4 (C-4), 70.2 (C-8), 73.3 (C-6), 126.1, 126.4, 126.5, 126.9, 128.1, 128.3, 128.4 (10 C–Ar), 170.7 (6C, 6 CO). ESI-MS calcd for C₂₉H₃₄N₄O₁₁ [M + Na]⁺ 637.33; found *m/z* 637.20.

Methyl (2,3-Dichlorobenzyl 5-Acetamido-4,7,8-tri-O-acetyl-9-azido-3,5,9-trideoxy- α -D-galacto-2-nonulopyranosid)onate (11e). Compound **10** (45.0 mg, 0.09 mmol) was reacted with 2,3-dichlorobenzyl alcohol (24.0 mg, 0.13 mmol), *N*-iodosuccinimide (24.0 mg, 0.12 mmol), and triflic acid (3.00 μ L, 6.00 mg, 0.04 mmol). The crude product was purified by chromatography on silica gel (0.5% gradient *i*PrOH in petrol ether/DCM 8:4) to yield **11e** (30.0 mg, 54%) as a yellow oil. ¹H NMR (500 MHz, CDCl₃) δ 1.83 (s, 3H, NHAc), 2.01 (t, *J* = 12.9 Hz, 1H, H-3a), 1.97, 2.08, 2.11 (3 s, 9H, 3 OAc), 2.63 (dd, *J* = 4.7, 12.9 Hz, 1H, H-3b), 3.19 (dd, *J* = 6.0, 13.6 Hz, 1H, H-9a), 3.50 (dd, *J* = 3.0, 13.6 Hz, 1H, H-9b), 3.72 (s, 3H, OMe), 4.05 (m, 2H, H-5, H-6), 4.50, 4.82 (A, B of AB, *J* = 13.0 Hz, 2H, CH₂Ar), 4.85 (m, 1H, H-4), 5.13 (d, *J* = 8.0 Hz, 1H, NHAc), 5.24 (m, 2H, H-7, H-8), 7.14 (t, *J* = 8.0 Hz, 1H, CH_{ar}), 7.34 (d, *J* = 8.1 Hz, 2H, CH_{ar}). ¹³C NMR (CDCl₃) δ 21.3, 21.5 (3C, 3 OAc), 23.6 (NHAc), 38.2 (C-3), 49.9 (C-5), 51.2 (C-9), 53.4 (OMe), 64.7 (CH₂Ar), 68.4

(C-8), 69.3 (C-4), 70.5 (C-7), 73.5 (C-6), 109.6 (C-2), 127.6, 127.7, 130.0 (6C, 6 C–Ar), 168.5, 170.5, 170.6, 170.7, 171.4 (5 CO). ESI-MS calcd for C₂₅H₃₀Cl₂N₄O₁₁ [M + Na]⁺ 655.13; found *m/z* 655.07.

Methyl (2,3-Difluorobenzyl 5-Acetamido-4,7,8-tri-O-acetyl-9-azido-3,5,9-trideoxy- α -D-galacto-2-nonulopyranosid)onate (11f). Compound **10** (86.0 mg, 0.17 mmol) was reacted with 2,3-difluorobenzyl alcohol (29.0 μ L, 37.0 mg, 0.26 mmol), *N*-iodosuccinimide (46.0 mg, 0.21 mmol), and triflic acid (6.00 μ L, 10.0 mg, 0.07 mmol). The crude product was purified by chromatography on silica gel (0.5% gradient *i*PrOH in petrol ether/DCM 8:4) to yield **11f** (67.0 mg, 66%) as a yellow oil. ¹H NMR (500 MHz, CDCl₃) δ 1.82 (s, 3H, NHAc), 1.95 (m, 1H, H-3a), 1.96, 2.10, 2.13 (3 s, 9H, 3 OAc), 2.58 (m, 1H, H-3b), 3.20 (m, 1H, H-9a), 3.50 (m, 1H, H-9b), 3.72 (s, 3H, OMe), 4.06 (m, 2H, H-5, H-6), 4.46, 4.75 (A, B of AB, *J* = 12.0 Hz, 2H, CH₂Ar), 4.82 (m, 1H, H-4), 5.29 (m, 2H, H-8, H-7), 7.05 (m, 3H, CH_{ar}). ¹³C NMR (CDCl₃) δ 21.3, 21.6 (3C, 3 OAc), 23.6 (NHAc), 38.2 (C-3), 49.8 (C-5), 51.3 (C-9), 53.3 (OMe), 60.8 (CH₂Ar), 68.4 (C-7), 69.3 (C-8), 70.3 (C-4), 73.4 (C-6), 99.1 (C-2), 117.4, 119.1 (*J* = 17.0 Hz), 124.3, 125.5 (6 C–Ar), 168.5, 170.5, 170.6, 170.8, 171.4 (5 CO). ESI-MS calcd for C₂₅H₃₀F₂N₄O₁₁ [M + 2Na]⁺ 645.99; found *m/z* 645.26.

Methyl (4-Chlorobenzyl 5-Acetamido-4,7,8-tri-O-acetyl-9-(4-chlorobenzamido)-3,5,9-trideoxy- α -D-galacto-2-nonulopyranosid)onate (12a). Compound **11a** (55.0 mg, 0.09 mmol) was reacted with *p*-chlorobenzoyl chloride (46.0 μ L, 63.0 mg, 0.36 mmol) and triphenylphosphine (47.0 mg, 0.18 mmol). The crude product was purified by chromatography on silica gel (0.5% gradient of MeOH in DCM) to yield **12a** (35.0 mg, 59%) as a yellow solid. [α]_D²⁰ −0.01 (*c* 2.9, CH₂Cl₂). ¹H NMR (500 MHz, CDCl₃) δ 1.89 (s, 3H, NHAc), 2.05 (m, 4H, H-3a, OAc), 2.14, 2.27 (2 s, 6H, 2 OAc), 2.67 (dd, *J* = 4.6, 12.8 Hz, 1H, H-3b), 2.92 (dt, *J* = 3.3, 14.9 Hz, 1H, H-9a), 3.65 (s, 3H, OMe), 4.05 (dd, *J* = 2.0, 10.7 Hz, 1H, H-6), 4.21 (q, *J* = 10.4 Hz, 1H, H-5), 4.36 (ddd, *J* = 3.0, 8.7, 15.1 Hz, 1H, H-9b), 4.41, 4.78 (A, B of AB, *J* = 12.3 Hz, 2H, CH₂Ar), 4.84 (m, 1H, H-4), 5.15 (dd, *J* = 2.0, 10.0 Hz, 1H, H-7), 5.31 (m, 2H, NHAc, H-8), 7.10 (dd, *J* = 4.0, 8.5 Hz, 1H, NH), 7.29 (m, 4H, CH_{ar}), 7.41, 7.77 (AA', BB' of AA'/BB', *J* = 8.5 Hz, 4H, CH_{ar}). ¹³C NMR (CDCl₃) δ 20.9, 21.2, 21.3 (3 OAc), 23.2 (NHAc), 38.1 (C-3), 38.4 (C-9), 49.5 (C-5), 52.7 (OMe), 66.1 (CH₂Ar), 67.9 (C-7), 68.2 (C-8), 69.0 (C-4), 72.2 (C-6), 98.4 (C-2), 128.4, 128.5, 128.8, 129.2, 132.0, 132.6, 133.5, 135.7, 137.7 (12 C–Ar), 168.1, 170.4, 171.2, 172.6 (6C, 6 CO). ESI-MS calcd for C₃₂H₃₆Cl₂N₂O₁₂ [M + Na]⁺ 733.16; found *m/z* 733.25.

Methyl (4-Fluorobenzyl 5-Acetamido-4,7,8-tri-O-acetyl-9-(4-chlorobenzamido)-3,5,9-trideoxy- α -D-galacto-2-nonulopyranosid)onate (12b). Compound **11b** (52.0 mg, 0.09 mmol) was reacted with *p*-chlorobenzoyl chloride (46.0 μ L, 63.0 mg, 0.36 mmol) and triphenylphosphine (52.0 mg, 0.19 mmol). The crude product was purified by chromatography on silica gel (0.5% gradient of MeOH in DCM) to yield **12b** (37.0 mg, 60%) as a yellow solid. ¹H NMR (500 MHz, CDCl₃) δ 1.85 (s, 3H, NHAc), 2.00 (m, 1H, H-3a), 2.03, 2.11, 2.26 (3 s, 9H, 3 OAc), 2.68 (dd, *J* = 4.5, 12.7 Hz, 1H, H-3b), 2.96 (dt, *J* = 3.5, 15.0 Hz, 1H, H-9a), 3.61 (s, 3H, OMe), 4.12 (d, *J* = 10.7 Hz, 1H, H-6), 4.23 (q, *J* = 10.2 Hz, 1H, H-5), 4.36 (ddd, *J* = 3.0, 8.7, 15.3 Hz, 1H, H-9b), 4.40, 4.78 (A, B of AB, *J* = 11.9 Hz, 2H, CH₂Ar), 4.87 (m, 1H, H-4), 5.20 (dd, *J* = 1.1, 9.9 Hz, 1H, H-7), 5.35 (dt, *J* = 2.9, 9.9 Hz, 1H, H-8), 5.99 (m, 1H, NHAc), 7.00 (t, *J* = 8.7 Hz, 2H, CH_{ar}), 7.17 (dd, *J* = 3.8 Hz, 8.1 Hz, 1H, NH), 7.49 (dd, *J* = 5.4 Hz, 8.5 Hz, 1H, CH_{ar}), 7.38, 7.76 (AA', BB' of AA'/BB', *J* = 8.4 Hz, 4H, CH_{ar}). ¹³C NMR (CDCl₃) δ 21.2, 21.3 (3 OAc), 23.1 (NHAc), 38.3 (C-3), 38.5 (C-9), 49.4 (C-5), 52.6 (OMe), 66.1 (CH₂Ar), 68.0 (C-7), 68.2 (C-8), 69.2 (C-4), 72.2 (C-6), 98.4 (C-2), 115.0, 115.2 (*J* = 21.3 Hz), 128.6, 129.6, 129.7 (*J* = 7.5 Hz), 132.0, 132.8, 132.9, 137.7, 161.3, 163.5 (*J* = 275.1 Hz), 168.1, 166.3 (*J* = 237.5 Hz) (12 C–Ar), 170.39, 170.4, 171.0, 172.5 (6C, 6 CO). ESI-MS calcd for C₃₂H₃₆ClFN₂O₁₂ [M + Na]⁺ 717.19; found *m/z* 717.34.

Methyl (Pentafluorobenzyl 5-Acetamido-4,7,8-tri-O-acetyl-9-(4-chlorobenzamido)-3,5,9-trideoxy-D-glycero- α -D-galacto-2-nonulopyranosid)onate (12c). Compound **11c** (47.0 mg, 0.07 mmol) was reacted with *p*-chlorobenzoyl chloride (36.0 μ L, 49.0 mg, 0.28 mmol) and triphenylphosphine (41.0 mg, 0.16 mmol). The crude product was purified by chromatography on silica gel (0.5% gradient of MeOH in DCM) to yield **12c** (25.0 mg, 45%). ¹H NMR (500 MHz, CDCl₃) δ 1.86 (s, 3H, NHAc), 1.94 (m, 1H, H-3a), 2.02, 2.16, 2.19 (3 s, 9H, 3 OAc), 2.60 (dd, *J* = 4.5, 12.8 Hz, 1H, H-3b), 3.55 (dd, *J* = 6.3, 12.3 Hz, 1H, H-9a), 3.81 (m, 4H, OMe, H-9b), 4.07 (d, *J* = 10.3 Hz, 1H, H-6), 4.20 (dd, *J* = 2.0, 10.8 Hz, 1H, H-5), 4.45, 4.93 (A, B of AB, *J* = 11.3 Hz, 2H, CH₂Ar), 5.37 (dd, *J* = 2.1, 8.1 Hz, 1H, H-7), 5.46 (m, 1H, H-8), 5.68 (d, *J* = 9.6 Hz, 1H, NHAc), 7.38, 8.00 (AA', BB' of AA'/BB', *J* = 8.4 Hz, 4H, CH_{ar}). ¹³C NMR (CDCl₃) δ 21.0, 21.2, 21.3 (3 OAc), 23.3 (NHAc), 37.8 (C-3), 43.7 (C-9), 49.5 (C-5), 53.1 (OMe), 54.1 (CH₂Ar), 68.3, 69.0 (C-4, C-7), 70.0 (C-8), 72.9 (C-6), 98.8 (C-2), 128.8, 128.9, 129.1, 131.5, 132.9, 133.2, 133.3 (12 C-Ar), 167.6, 167.9, 168.7, 170.3, 170.4, 170.6 (6 CO). ESI-MS calcd for C₃₂H₃₂ClF₅N₂O₁₂ [M + Cl]⁻ 801.45; found *m/z* 801.36.

Methyl (2-Naphthyl 5-Acetamido-4,7,8-tri-O-acetyl-9-(4-chlorobenzamido)-3,5,9-trideoxy-D-glycero- α -D-galacto-2-nonulopyranosid)onate (12d). Compound **11d** (50.0 mg, 0.08 mmol) was reacted with *p*-chlorobenzoyl chloride (42.0 μ L, 57.0 mg, 0.32 mmol) and triphenylphosphine (47.0 mg, 0.18 mmol). The crude product was purified by chromatography on silica gel (0.5% gradient of MeOH in DCM) to yield **12d** (25.0 mg, 47%). ¹H NMR (500 MHz, CDCl₃) δ 1.87 (s, 3H, NHAc), 2.04 (s, 3H, OAc), 2.10 (t, *J* = 12.4 Hz, 1H, H-3a), 2.14, 2.24 (2 s, 6H, 2 OAc), 2.72 (dd, *J* = 4.5, 12.8 Hz, 1H, H-3b), 2.92 (m, 1H, H-9a), 3.59 (s, 3H, OMe), 4.11 (m, 1H, H-6), 4.23 (q, *J* = 10.3 Hz, 1H, H-5), 4.34 (ddd, *J* = 3.0, 8.5, 15.2 Hz, 1H, H-9b), 4.63 (A of AB, *J* = 12.3 Hz, 1H, CH₂Ar), 4.87 (dt, *J* = 4.6, 12.3 Hz, 1H, H-4), 4.98 (B of AB, *J* = 12.3 Hz, 1H, CH₂Ar), 5.17 (d, *J* = 9.8 Hz, 1H, H-7), 5.34 (d, *J* = 9.9 Hz, 1H, H-8), 5.50 (m, 1H, NHAc), 7.11 (m, 1H, NH), 7.39 (AA' of AA'/BB', *J* = 8.5 Hz, 2H, CH_{ar}), 7.44–7.47 (m, 3H, CH_{ar}), 7.76 (BB' of AA'/BB', *J* = 8.5 Hz, 2H, CH_{ar}), 7.75–7.83 (m, 4H, CH_{ar}). ¹³C NMR (CDCl₃) δ 21.3, 21.5, 21.6 (3 OAc), 23.5 (NHAc), 38.6 (C-3), 38.9 (C-9), 49.9 (C-5), 53.0 (OMe), 67.4 (CH₂Ar), 68.4 (C-7), 68.7 (C-8), 69.5 (C-4), 72.6 (C-6), 98.9 (C-2), 126.2, 126.3, 126.5, 127.0, 128.0, 128.3, 128.9, 129.0, 129.2, 132.4, 133.1, 133.3, 133.5, 135.1 (14 C-Ar), 166.6, 168.7, 170.8, 171.5, 172.9 (6C, 6 CO). ESI-MS calcd for C₃₆H₃₉ClN₂O₁₂ [M + Na]⁺ 749.22; found *m/z* 749.25.

Methyl (2,3-Dichlorobenzyl 5-Acetamido-4,7,8-tri-O-acetyl-9-(4-chlorobenzamido)-3,5,9-trideoxy-D-glycero- α -D-galacto-2-nonulopyranosid)onate (12e). Compound **11e** (47.0 mg, 0.07 mmol) was reacted with *p*-chlorobenzoyl chloride (38.0 μ L, 52.0 mg, 0.30 mmol) and triphenylphosphine (43.0 mg, 0.16 mmol). The crude product was purified by chromatography on silica gel (0.5% gradient of MeOH in DCM) to yield **12e** (25.0 mg, 47%) as a yellow solid. ¹H NMR (500 MHz, CDCl₃) δ 1.78 (t, *J* = 12.4 Hz, 1H, H-3a), 1.80 (s, 3H, NHAc), 1.97, 2.04, 2.17 (3 s, 9H, 3 OAc), 2.63 (dd, *J* = 4.8, 12.8 Hz, 1H, H-3b), 2.90 (ddd, *J* = 3.3, 3.7, 15.1 Hz, 1H, H-9a), 3.64 (s, 3H, OMe), 4.01 (d, *J* = 10.8 Hz, 1H, H-6), 4.15 (q, *J* = 10.4 Hz, 1H, H-5), 4.23 (m, 1H, H-9b), 4.49 (A of AB, *J* = 13.0 Hz, 1H, CH₂Ar), 4.30 (m, 2H, H-4, CH₂Ar), 5.11 (d, *J* = 1.6 Hz, 1H, H-7), 5.20 (m, 1H, H-8), 5.55 (d, *J* = 8.8 Hz, 1H, NHAc), 7.03 (dd, *J* = 3.8 Hz, 1H, NH), 7.13 (t, *J* = 7.8 Hz, 1H, CH_{ar}), 7.39 (m, 2H, CH_{ar}), 7.68, 7.93 (AA', BB' of AA'/BB', *J* = 8.5 Hz, 4H, CH_{ar}). ¹³C NMR (CDCl₃) δ 21.3, 21.5, 21.6 (3 OAc), 23.5 (NHAc), 38.2 (C-3), 39.0 (C-9), 49.9 (C-5), 53.2 (OMe), 64.8 (CH₂Ar), 68.4 (C-8), 68.8 (C-7), 69.5 (C-4), 72.7 (C-6), 110.0 (C-2), 127.6, 128.1, 130.0, 131.8, 133.0, 136.0, 138.1 (12 C-Ar), 160.0, 168.0, 170.6, 170.9, 171.5, 172.8 (6 CO). ESI-MS calcd for C₃₂H₃₂Cl₂N₂O₁₂ [M + Na]⁺ 767.13; found *m/z* 767.12.

Methyl (2,3-Difluorobenzyl 5-Acetamido-4,7,8-tri-O-acetyl-9-(4-chlorobenzamido)-3,5,9-trideoxy-D-glycero- α -D-galacto-2-nonulopyranosid)onate (12f). Compound **11f** (67.0 mg, 0.11 mmol)

was reacted with *p*-chlorobenzoyl chloride (57.0 μ L, 78.0 mg, 0.45 mmol) and triphenylphosphine (65.0 mg, 0.25 mmol). The crude product was purified by chromatography on silica gel (0.5% gradient of MeOH in DCM) to yield **12f** (44.0 mg, 55%) as a yellow solid. ¹H NMR (500 MHz, CDCl₃) δ 1.80 (s, 3H, NHAc), 2.04 (m, 4H, H-3a, OAc), 2.14, 2.27 (2 s, 6H, 2 OAc), 2.65 (dd, *J* = 4.6, 12.8 Hz, 1H, H-3b), 2.97 (dt, *J* = 3.5, 15.0 Hz, 1H, H-9a), 3.74 (s, 3H, OMe), 4.08 (m, 1H, H-6), 4.23 (q, *J* = 10.4 Hz, 1H, H-5), 4.33 (m, 1H, H-9b), 4.52, 4.84 (A, B of AB, *J* = 12.0 Hz, 2H, CH₂Ar), 5.17 (dd, *J* = 1.9, 9.8 Hz, 1H, H-7), 5.31 (m, 1H, H-8), 5.43 (m, 1H, NHAc), 7.75 (m, 4H, NH, CH_{ar}), 7.39, 7.77 (AA', BB' of AA'/BB', *J* = 8.5 Hz, 4H, CH_{ar}). ¹³C NMR (CDCl₃) δ 21.3, 21.5, 21.6 (3 OAc), 23.5 (NHAc), 38.3 (C-3), 38.9 (C-9), 49.9 (C-5), 53.2 (OMe), 60.8 (CH₂Ar), 68.3 (C-8), 68.7 (C-7), 69.4 (C-4), 72.7 (C-6), 98.0 (C-2), 117.3, 117.4 (*J* = 17.0 Hz), 125.7, 128.8, 129.2, 132.4, 132.5 (*J* = 9.0 Hz), 133.0, 138.1 (12 C-Ar), 166.7, 168.4, 170.7, 170.9, 171.6, 172.9 (6 CO). ESI-MS calcd for C₃₂H₃₂ClF₂N₂O₁₂ [M + Na]⁺ 735.18; found *m/z* 735.15.

Sodium (4-Chlorobenzyl 5-Acetamido-9-(4-chlorobenzamido)-3,5,9-trideoxy-D-glycero- α -D-galacto-2-nonulopyranosid)onate (13a). Compound **12a** (25.0 mg, 0.03 mmol) was treated with 10% aqueous NaOH (0.2 mL) in MeOH (1.0 mL). The crude product was purified by chromatography on RP-18 (5% gradient of MeOH in H₂O) to yield **13a** as a white solid (15.0 mg, 50%). [α]_D²⁰ = -0.25 (c 0.41, H₂O). ¹H NMR (500 MHz, CD₃OD) δ 1.55 (dd, *J* = 3.3, 11.9 Hz, 1H, H-3a), 1.90 (s, 3H, NHAc), 2.81 (dd, *J* = 2.0, 12.2 Hz, 1H, H-3b), 3.34 (dd, *J* = 1.8, 9.0 Hz, 1H, H-7), 3.41 (dd, *J* = 7.8, 13.6 Hz, 1H, H-9a), 3.55–3.63 (m, 3H, H-4, H-5, H-6), 3.67 (dd, *J* = 3.1 Hz, 13.6 Hz, 1H, H-9b), 3.94 (m, 1H, H-8), 4.40, 4.70 (A, B of AB, *J* = 11.6 Hz, 2H, CH₂Ar), 7.18, 7.25 (AA', BB' of AA'/BB', *J* = 8.4 Hz, 4H, CH_{ar}), 7.36, 7.72 (AA', BB' of AA'/BB', *J* = 8.6 Hz, 4H, CH_{ar}). ¹³C NMR (CD₃OD) δ 22.6 (NHAc), 42.7 (C-3), 44.6 (C-9), 54.2 (C-5), 66.5 (CH₂Ar), 69.6 (C-4), 71.2 (C-8), 72.5 (C-7), 74.4 (C-6), 102.1 (C-2), 129.2, 129.7, 130.1, 130.6, 134.0, 134.5, 138.6, 138.8 (12 C-Ar), 169.2, 174.3, 175.5 (3 CO). HRMS calcd for C₂₅H₂₈Cl₂N₂O₉ [M - H]⁻ 615.0891; found 615.0888.

Sodium (4-Fluorobenzyl 5-Acetamido-9-(4-chlorobenzamido)-3,5,9-trideoxy-D-glycero- α -D-galacto-2-nonulopyranosid)onate (13b). Compound **12b** (37.0 mg, 0.05 mmol) was treated with 10% aqueous NaOH (0.3 mL) in MeOH (1.2 mL). The crude product was purified by chromatography on RP-18 (5% gradient of MeOH in H₂O) to yield **13b** as a white solid (21.1 mg, 70%). [α]_D²⁰ = -0.22 (c 0.33, H₂O). ¹H NMR (500 MHz, CD₃OD) δ 1.54 (m, 1H, H-3a), 1.90 (s, 3H, NHAc), 2.80 (m, 1H, H-3b), 3.35 (m, 1H, H-7), 3.42 (m, 1H, H-9a), 3.58–3.69 (m, 4H, H-4, H-5, H-6, H-9b), 3.96 (m, 1H, H-8), 4.39, 4.69 (A, B of AB, *J* = 11.2 Hz, 2H, CH₂Ar), 6.90 (t, *J* = 8.8 Hz, 2H, CH_{ar}), 7.27 (m, 2H, CH_{ar}), 7.35 (m, 2H, CH_{ar}), 7.72 (BB' of AA'/BB', *J* = 8.5 Hz, 2H, CH_{ar}). ¹³C NMR (CD₃OD) δ 22.7 (NHAc), 42.7 (C-3), 44.5 (C-9), 54.2 (C-5), 66.7 (CH₂Ar), 69.6 (C-4), 71.4 (C-8), 72.5 (C-7), 74.4 (C-6), 102.1 (C-2), 115.8 (*J* = 12.5 Hz), 130.1, 129.7, 131.1 (*J* = 21.3 Hz), 138.6, 162.0 (12 C-Ar), 164.6, 169.3, 174.4 (3 CO). HRMS calcd for C₂₅H₂₈ClFN₂O₉ [M - H]⁻ 577.1367; found 577.1369.

Sodium (Pentafluorobenzyl 5-Acetamido-9-(4-chlorobenzamido)-3,5,9-trideoxy-D-glycero- α -D-galacto-2-nonulopyranosid)onate (13c). Compound **12c** (25.0 mg, 0.03 mmol) was treated with 10% aqueous NaOH (0.2 mL) in MeOH (1.0 mL). The crude product was purified by chromatography on RP-18 (5% gradient of MeOH in H₂O) to yield **13c** as a white solid (8.0 mg, 39%). [α]_D²⁰ = -0.02 (c 0.32, H₂O). ¹H NMR (500 MHz, D₂O) δ 1.63 (t, *J* = 12.1 Hz, 1H, H-3a), 1.97 (s, 3H, NHAc), 2.71 (dd, *J* = 4.7, 12.4 Hz, 1H, H-3b), 3.50 (m, 2H, H-7, H-9a), 3.65 (ddd, *J* = 4.7, 9.5, 11.9 Hz, 1H, H-4), 3.74–3.79 (m, 3H, H-5, H-6, H-9b), 3.88 (ddd, *J* = 2.9, 7.8, 8.8 Hz, 1H, H-8), 4.66, 4.86 (A, B of AB, *J* = 11.7 Hz, 2H, CH₂Ar), 7.50, 7.73 (AA', BB' of AA'/BB', *J* = 8.6 Hz, 4H, CH_{ar}). ¹³C NMR (D₂O) δ 21.9 (NHAc), 40.3 (C-3), 42.6 (C-9), 51.8 (C-5), 54.1 (CH₂Ar), 68.1 (C-4), 69.8 (C-7), 70.3 (C-8), 72.8 (C-6), 101.0 (C-2), 128.7, 128.7, 132.2, 136.2, 137.5, 139.1

(12 C-Ar), 170.1, 172.8, 175.0 (3 CO). HRMS calcd for $C_{25}H_{23}ClF_3N_2NaO_9 [M - H]^-$ 625.1018; found 625.1015.

Sodium (2-Naphthyl 5-Acetamido-9-(4-chlorobenzamido)-3,5,9-trideoxy- α -D-galacto-2-nonulopyranosid)onate (13d). Compound **12d** (25.0 mg, 0.03 mmol) was treated with 10% aqueous NaOH (0.2 mL) in MeOH (1.0 mL). The crude product was purified by chromatography on RP-18 (5% gradient of MeOH in H_2O) to yield **13d** as a white solid (14.0 mg, 70%). $[\alpha]_D^{20} -0.37$ (*c* 0.46, H_2O). 1H NMR (500 MHz, CD_3OD) δ 1.60 (m, 1H, H-3a), 1.90 (s, 3H, NHAc), 2.84 (d, *J* = 11.8 Hz, H-3b), 3.35 (d, *J* = 8.9 Hz, H-7), 3.42 (dd, *J* = 7.7, 13.7 Hz, H-9a), 3.60–3.68 (m, 4H, H-4, H-5, H-6, H-9b), 4.60 (A of AB, *J* = 11.5 Hz, 1H, CH_2Ar), 3.96 (m, 1H, H-8), 4.89 (B of AB, *J* = 11.5 Hz, 1H, CH_2Ar), 7.31–7.36 (m, 4H, CH_{ar}), 7.40 (dd, *J* = 1.5, 8.4 Hz, 1H, CH_{ar}), 7.67–7.72 (m, 6H, CH_{ar}). ^{13}C NMR (CD_3OD) δ 21.7 (NHAc), 43.3 (C-3), 43.4 (C-9), 53.1 (C-5), 68.6 (CH_2Ar), 71.3 (C-7), 72.0 (C-4), 73.3 (C-6), 110.0 (C-2), 125.7, 125.9, 126.3, 126.4, 127.6, 127.7, 127.9, 128.6, 129.1 (16 C-Ar), 174.6, 175.3 (3C, CO). HRMS calcd for $C_{29}H_{30}ClN_2NaO_9 [M + Na]^+$ 631.1438; found 631.1435.

Sodium (2,3-Dichlorobenzyl 5-Acetamido-9-(4-chlorobenzamido)-3,5,9-trideoxy- α -D-galacto-2-nonulopyranosid)onate (13e). Compound **12e** (25.0 mg, 0.03 mmol) was treated with 10% aqueous NaOH (0.2 mL) in MeOH (1.0 mL). The crude product was purified by chromatography on RP-18 (5% gradient of MeOH in H_2O) to yield **13e** as a white solid (10.0 mg, 50%). $[\alpha]_D^{20} -0.19$ (*c* 0.53, H_2O). 1H NMR (500 MHz, CD_3OD) δ 1.69 (t, *J* = 11.8 Hz, 1H, H-3a), 1.93 (s, 3H, NHAc), 2.01 (dd, *J* = 9.1, 11.8 Hz, 1H, H-3b), 3.36 (dd, *J* = 1.7, 9.0 Hz, 1H, H-7), 3.45 (dd, *J* = 7.6, 13.6 Hz, 1H, H-9a), 3.59 (m, 1H, H-6), 3.65–3.70 (m, 3H, H-4, H-5, H-9b), 3.92 (m, 1H, H-8), 4.65, 4.89 (A, B of AB, *J* = 13.6 Hz, 2H, CH_2Ar), 7.18 (t, *J* = 7.9 Hz, 1H, CH_{ar}), 7.32 (dd, *J* = 1.5, 8.0 Hz, 1H, CH_{ar}), 7.39 (AA' of AA'/BB', *J* = 8.5 Hz, 2H, CH_{ar}), 7.49 (dd, *J* = 1.1, 7.7 Hz, 1H, CH_{ar}), 7.74 (BB' of AA'/BB', *J* = 8.5 Hz, 2H, CH_{ar}). ^{13}C NMR (CD_3OD) δ 21.6 (NHAc), 41.5 (C-3), 43.4 (C-9), 53.1 (C-5), 63.5 (CH_2Ar), 68.6 (C-4), 70.2 (C-8), 71.5 (C-7), 73.4 (C-6), 102.5 (C-2), 113.3, 117.0, 118.4, 127.4, 127.5, 128.6, 128.9, 129.0, 139.6 (12 C-Ar), 163.8, 174.5 (3C, 3 CO). HRMS calcd for $C_{25}H_{27}Cl_2N_2O_9 [M + Na]^+$ 627.0682; found 627.0683.

Sodium (2,3-Difluorobenzyl 5-Acetamido-9-(4-chlorobenzamido)-3,5,9-trideoxy- α -D-galacto-2-nonulopyranosid)onate (13f). Compound **12f** (44.0 mg, 0.06 mmol) was treated with 10% aqueous NaOH (0.3 mL) in MeOH (1.5 mL). The crude product was purified by chromatography on RP-18 (5% gradient of MeOH in H_2O) to yield **13f** as a white solid (23.0 mg, 64%). $[\alpha]_D^{20} -0.18$ (*c* 1.13, H_2O). 1H NMR (500 MHz, CD_3OD) δ 1.56 (m, 1H, H-3a), 1.90 (s, 3H, NHAc), 2.84 (dd, *J* = 1.1, 7.7 Hz, 1H, H-3b), 3.37 (d, *J* = 9.0 Hz, 1H, H-7), 3.46 (dd, *J* = 7.7, 13.6 Hz, 1H, H-9a), 3.59 (m, 1H, H-6), 3.63–3.70 (m, 3H, H-4, H-5, H-9b), 3.96 (m, 1H, H-8), 4.60, 4.85 (A, B of AB, *J* = 12.2 Hz, 2H, CH_2Ar), 7.04 (m, 2H, CH_{ar}), 7.24 (t, *J* = 6.5 Hz, 1H, CH_{ar}), 7.38, 7.75 (AA', BB' of AA'/BB', *J* = 8.5 Hz, 4H, CH_{ar}). ^{13}C NMR (CD_3OD) δ 21.6 (NHAc), 41.5 (C-3), 43.5 (C-9), 53.1 (C-5), 59.2 (CH_2Ar), 68.6 (C-4), 70.3 (C-8), 71.4 (C-7), 73.4 (C-6), 101.0 (C-2), 115.9, 116.1 (*J* = 17.3 Hz), 124.2, 125.3, 128.6, 129.1, 133.5, 137.6 (12 C-Ar), 168.3, 173.0, 174.5 (3 CO). HRMS calcd for $C_{25}H_{27}ClF_2N_2O_9 [M + Na]^+$ 595.1273; found 595.1272.

Methyl (Methyl 5-tert-Butyloxycarbonylamino-4,7,8-tri-O-acetyl-9-azido-3,5,9-trideoxy-2-thio- α -D-galacto-2-nonulopyranosid)onate (14). Compound **10** (85.0 mg, 0.17 mmol) was dissolved in dry THF (0.7 mL) under argon. Boc₂O (74.0 mg, 0.34 mmol) was added to the reaction mixture, followed by DMAP (4.50 mg, 0.02 mmol). The reaction mixture was heated up to 60 °C for 5 h. After cooling to rt, MeOH (0.7 mL) and $N_2H_4 \cdot H_2O$ (52 μ L, 1.1 mmol) were added and stirring was continued for 16 h. The reaction mixture was washed successively with 0.1 M HCl (1 \times 5 mL), 0.5 M $CuSO_4$ (1 \times 5 mL), saturated aqueous $NaHCO_3$ (2 \times 5 mL), and H_2O (1 \times 5 mL). The organic layer was dried over Na_2SO_4 , filtered, and concentrated under

reduced pressure to give a yellow oil. The crude product was reacted with acetic anhydride (0.9 mL) in dry pyridine (1.7 mL). A catalytic amount of DMAP was added, and stirring was continued at rt. After 18 h, the reaction mixture was washed with $CuSO_4$ (0.5 M, 4 \times 5 mL), saturated aqueous $NaHCO_3$ (1 \times 5 mL), and H_2O (1 \times 5 mL). The organic layer was dried over Na_2SO_4 , filtered, and the solvent was evaporated. The crude product was purified by chromatography on silica gel (EA:PE, gradient 1:1 to 2:1) to yield **14** (72 mg, 76%) as a white foam. $[\alpha]_D^{20} 0.31$ (*c* 1.68, CH_2Cl_2). 1H NMR (500 MHz, $CDCl_3$) δ 1.37 (s, 9H, *t*-Butyl), 1.95 (m, 1H, H-3a), 2.02 (s, 3H, SMe), 2.10, 2.14, 2.17 (3 s, 9H, 3 OAc), 2.73 (dd, *J* = 4.6, 12.7 Hz, 1H, H-3b), 3.25 (dd, *J* = 6.0, 13.5 Hz, 1H, H-9a), 3.61 (dd, *J* = 3.2, 13.5 Hz, 1H, H-9b), 3.73 (m, 1H, H-6), 3.80 (m, 4H, OMe, H-5), 4.25 (d, *J* = 10.4 Hz, 1H, NH), 4.78 (m, 1H, H-4), 5.29 (m, 1H, H-8), 5.41 (m, 1H, H-7). ^{13}C NMR ($CDCl_3$) δ 12.1 (SMe), 20.8, 21.0, 21.3 (3 OAc), 27.9 (3C, C- $(CH_3)_3$), 38.0 (C-3), 50.4 (C-9), 50.6 (C-5), 52.9 (OMe), 67.6 (C-7), 68.1 (C-8), 70.0 (C-4), 73.7 (C-6), 80.2 (C-2), 83.0 (C- $(CH_3)_3$), 155.2 (CONH), 168.2, 170.0, 170.6, 174.0 (4 CO). ESI-MS calcd for $C_{22}H_{34}N_4O_{11}S [M + Na]^+$ 585.19; found *m/z* 585.15.

Methyl (Methyl 5-Amino-4,7,8-tri-O-acetyl-9-azido-3,5,9-trideoxy-2-thio- α -D-galacto-2-nonulopyranosid)onate (15). Compound **14** (120 mg, 0.21 mmol) was dissolved in 4 M PhOH (in abs DCM; 7.5 mL) and 4 M TMSCl (in abs DCM; 1.5 mL). The reaction mixture was stirred at rt for 2 h. The reaction mixture was washed with saturated aqueous $NaHCO_3$ (2 \times 5 mL) and H_2O (1 \times 5 mL). The organic layer was dried over Na_2SO_4 , filtered, and the solvent was evaporated. The crude product was purified by chromatography on silica gel (EA:PE, gradient 1:1 to 8:1) to yield **15** as a white foam (74 mg, 70%). 1H NMR (500 MHz, $CDCl_3$) δ 2.01 (s, 3H, SMe), 2.04 (m, 1H, H-3a), 2.10, 2.11, 2.19 (3 s, 9H, 3 OAc), 2.71 (dd, *J* = 4.7, 12.8 Hz, 1H, H-3b), 3.38 (d, *J* = 10.5 Hz, 1H, H-6), 3.66 (dd, *J* = 3.3, 13.3 Hz, 1H, H-9a), 3.70 (dd, *J* = 2.8, 13.3 Hz, 1H, H-9b), 3.76 (dd, *J* = 1.3, 9.4 Hz, 1H, H-7), 3.78 (s, 3H, OMe), 3.98 (dt, *J* = 8.0, 10.5 Hz, 1H, H-5), 4.93 (dd, *J* = 6.9, 10.7 Hz, 1H, H-4), 5.26 (t, *J* = 3.9, 9.4 Hz, 1H, H-8), 5.94 (d, *J* = 7.8 Hz, 2H, NH₂). ^{13}C NMR ($CDCl_3$) δ 12.0 (SMe), 21.1, 21.2, 23.2 (3 OAc), 37.7 (C-3), 51.2, 52.1, 52.9 (3C, C-5 C-9, OMe), 66.8, 69.3, 70.0 (3C, C-4, C-7, C-8), 75.3 (C-6), 82.2 (C-2), 167.9, 170.2, 172.4, 172.9 (4 CO). ESI-MS calcd for $C_{17}H_{26}N_4O_6S [M + Na]^+$ 485.14; found *m/z* 485.13.

Methyl (Methyl 5-Fluoroacetamido-4,7,8-tri-O-acetyl-9-azido-3,5,9-trideoxy-2-thio- α -D-galacto-2-nonulopyranosid)onate (16a). Compound **15** (74.0 mg, 0.16 mmol) was dissolved in dry DCM (1.8 mL) and cooled to 0 °C. Then, monofluoroacetic chloride was added dropwise (30.0 mg, 31.0 μ L, 0.32 mmol), followed by the addition of NEt_3 (324 mg, 0.45 mL, 3.2 mmol) and DMAP (9.8 mg, 0.08 mmol). Stirring was continued overnight, and the reaction mixture was allowed to come to rt. The brown solution was washed with saturated aqueous $NaHCO_3$ (3 \times 5 mL), saturated aqueous NaCl (1 \times 5 mL), and H_2O (1 \times 5 mL). The organic phase was dried over Na_2SO_4 , filtered, and concentrated under reduced pressure. The crude product was purified by chromatography on silica gel (EA:PE, gradient 1:1 to 2:1) to yield **16a** (71 mg, 85%). $[\alpha]_D^{20} 0.34$ (*c* 1.2, CH_2Cl_2). 1H NMR (500 MHz, $CDCl_3$) δ 1.98 (m, 1H, H-3a), 2.03 (s, 3H, SMe), 2.12, 2.15, 2.21 (3 s, 9H, 3 OAc), 2.79 (dd, *J* = 4.7, 12.8 Hz, 1H, H-3b), 3.24 (dd, *J* = 4.8, 13.6 Hz, 1H, H-9a), 3.64 (dd, *J* = 2.8, 13.6 Hz, 1H, H-9b), 3.84 (s, 3H, OMe), 3.90 (dd, *J* = 1.9, 10.7 Hz, 1H, H-6), 4.14 (m, 1H, H-5), 4.71 (m, 2H, CH_2F), 4.92 (td, *J* = 4.7, 11.6 Hz, 1H, H-4), 5.32 (m, 2H, H-7, H-8), 6.09 (dd, *J* = 3.3, 10.3 Hz, 1H, NH). ^{13}C NMR ($CDCl_3$) δ 12.2 (SMe), 20.9 (3C, OAc), 37.8 (C-3), 48.6 (C-5), 50.6 (C-9), 53.1 (OMe), 67.8 (C-7), 69.4 (C-6), 69.6 (2C, C-4, C-8), 74.1 (C-6), 80.8 (*J* = 186.1 Hz, CH_2F), 83.1 (C-2), 168.2, 170.2, 170.6 (5C, CO). ESI-MS calcd for $C_{19}H_{27}FN_4O_{10}S [M + Na]^+$ 545.14; found *m/z* 545.15.

Methyl (Methyl 5-Chloroacetamido-4,7,8-tri-O-acetyl-9-azido-3,5,9-trideoxy-2-thio- α -D-galacto-2-nonulopyranosid)onate (16b). Compound **15** (55.0 mg, 0.12 mmol) was dissolved in dioxane/water (0.5 mL/0.1 mL), treated with triethylamine

(48.6 mg, 34 μ L, 0.48 mmol), and cooled to 0 °C. Then, chloroacetic anhydride (41.0 mg, 0.48 mmol) was added, and stirring was continued at rt for 3 h. The reaction mixture was diluted with CHCl_3 (10.0 mL) and washed successively with saturated aqueous NaHCO_3 (3 \times 5 mL) and H_2O (1 \times 5 mL). The organic layer was dried over Na_2SO_4 , filtered, and concentrated under reduced pressure. The crude product was purified by chromatography on silica gel (gradient, PE:EA; 1:1 to 1:2) to yield **16b** as a white foam (42 mg, 66%). $[\alpha]_{\text{D}}^{20}$ 0.33 (c 0.95, CH_2Cl_2). ^1H NMR (500 MHz, CDCl_3) δ 1.91 (t, J = 12.2 Hz, 1H, H-3a), 1.96 (s, 3H, SMe), 2.06, 2.08, 2.13 (3 s, 9H, 3 OAc), 2.72 (dd, J = 4.7 Hz, 12.8 Hz, 1H, H-3b), 3.18 (m, 1H, H-9a), 3.59 (m, 1H, H-9b), 3.78 (s, 3H, OMe), 3.85 (A of AB, J = 15.0 Hz, 1H, CH_2Cl), 3.86 (m, 1H, H-6), 3.93 (B of AB, J = 15.0 Hz, 1H, CH_2Cl), 4.03 (m, 1H, H-5), 4.90 (dt, J = 4.7 Hz, 11.5 Hz, 1H, H-4), 5.24 (m, 2H, H-7, H-8), 6.28 (d, J = 10.1 Hz, 1H, NH). ^{13}C NMR (CDCl_3) δ 12.1 (SMe), 20.8, 20.9, 21.1 (3 OAc), 37.9 (C-3), 42.5 (CH_2Cl), 49.7 (C-9), 50.6 (C-5), 53.1 (OMe), 67.1, 67.9, 69.2 (3C, C-4, C-7, C-8), 74.2 (C-6), 83.1 (C-2), 166.6, 167.8, 170.3, 170.6 (5C, CO). ESI-MS calcd for $\text{C}_{19}\text{H}_{27}\text{ClN}_4\text{O}_{16}\text{S}[\text{M} + \text{Na}]^+$ 561.10; found m/z 561.37.

Methyl (Methyl 5-*o*-Nitrotoluenesulfonamido-4,7,8-tri-*O*-acetyl-9-azido-3,5,9-trideoxy-2-thio-*D*-glycero- α -*D*-galacto-2-nonulopyranosid)onate (16c). Compound **15** (73.0 mg, 0.16 mmol) was dissolved in dry DCM (3.0 mL), and it was cooled to 0 °C. Nosylchloride (105 mg, 0.47 mmol), NEt_3 (34.0 μ L, 48.0 mg, 0.47 mmol), and DMAP (10.0 mg, 0.08 mmol) were added successively. The reaction mixture was stirred at rt overnight. Then it was washed with saturated aqueous NaHCO_3 (2 \times 5 mL) and H_2O (1 \times 5 mL). The organic phase was dried over Na_2SO_4 , filtered, and concentrated under reduced pressure. The crude product was purified by chromatography on silica gel (EA:PE, gradient 1:1 to 2:1) to yield **16c** (81 mg, 78%). ^1H NMR (500 MHz, CDCl_3) δ 1.85 (m, 1H, H-3a), 2.02 (s, 3H, SMe), 2.10, 2.13, 2.21 (3 s, 9H, 3 OAc), 2.80 (m, 1H, H-3b), 3.32 (dd, J = 6.2, 13.4 Hz, 1H, H-9a), 3.57 (dd, J = 3.2, 13.4 Hz, 1H, H-9b), 3.80 (m, 1H, H-5), 3.82 (s, 3H, OMe), 3.91 (d, J = 10.5 Hz, 1H, H-6), 4.97 (td, J = 4.7, 11.4 Hz, 1H, H-4), 5.30 (m, 2H, H-7, H-8), 5.75 (d, J = 9.4 Hz, 1H, NH), 7.70 (m, 2H, CH_{ar}), 7.90 (d, J = 7.9, 1H, CH_{ar}), 8.10 (d, J = 6.5, 1H, CH_{ar}). ^{13}C NMR (CDCl_3) δ 12.1 (SMe), 20.5, 21.1, 21.1 (3 OAc), 38.1 (C-3), 50.8 (C-9), 53.2, 53.5 (2C, C-5, OMe), 68.8 (C-7), 69.7 (C-4), 70.4 (C-8), 74.7 (C-6), 82.8 (C-2), 125.5, 130.4, 133.5, 135.5, 147.5 (6C, C-Ar), 167.8, 170.4 (4C, CO). ESI-MS calcd for $\text{C}_{23}\text{H}_{29}\text{N}_5\text{O}_{13}\text{S}_2[\text{M} - \text{H}]^-$ 646.12; found m/z 646.56.

Methyl (Methyl 5-Methylsulfonamido-4,7,8-tri-*O*-acetyl-9-azido-3,5,9-trideoxy-2-thio-*D*-glycero- α -*D*-galacto-2-nonulopyranosid)onate (16d). Compound **15** (50.0 mg, 0.11 mmol) was dissolved in dry DCM (2.0 mL) under argon atmosphere and subsequently cooled to 0 °C. Methanesulfonylchloride (25.0 μ L, 37.0 mg, 0.32 mmol), NEt_3 (45.0 μ L, 33.0 mg, 0.32 mmol), and a catalytic amount of DMAP were added successively. The reaction mixture was stirred at 0 °C overnight. Then it was washed with saturated aqueous NaHCO_3 (2 \times 5 mL) and H_2O (1 \times 5 mL). The organic phase was dried over Na_2SO_4 , filtered, and concentrated under reduced pressure. The crude product was purified by chromatography on silica gel (EA:PE, gradient 1:1 to 2:1) to yield **16d** (30 mg, 66%). ^1H NMR (500 MHz, CDCl_3) δ 1.98 (t, J = 12.3 Hz, 1H, H-3a), 2.03 (s, 3H, SMe), 2.13, 2.17 (3 s, 9H, 3 OAc), 2.68 (dd, J = 4.7, 12.7 Hz, 1H, H-3b), 3.14 (s, 3H, CH_3), 3.29 (dd, J = 6.1, 13.4 Hz, 1H, H-9a), 3.66 (m, 2H, H-5, H-9b), 3.77 (s, 3H, OMe), 3.88 (dd, J = 1.6, 10.6 Hz, 1H, H-6), 4.55 (d, J = 9.8 Hz, 1H, NH), 5.16 (td, J = 4.7, 11.5 Hz, 1H, H-4), 5.26 (m, 1H, H-8), 5.48 (dd, J = 1.6, 7.2 Hz, 1H, H-7). ^{13}C NMR (CDCl_3) δ 14.3 (SMe), 20.9, 21.1, 21.4 (3 OAc), 31.7 (CH_3), 38.2 (C-3), 50.7 (C-9), 52.5 (C-5), 53.0 (OMe), 68.3, 69.6 (2C, C-7, C-8), 70.4 (C-4), 74.5 (C-6), 82.8 (C-2), 167.8, 170.3, 171.7, 172.2 (4 CO). ESI-MS calcd for $\text{C}_{18}\text{H}_{28}\text{N}_4\text{O}_{11}\text{S}_2[\text{M} + \text{Na}]^+$ 563.12; found m/z 563.18.

Methyl ((2,3-Difluorobenzyl) 5-Fluoroacetamido-4,7,8-tri-*O*-acetyl-9-azido-3,5,9-trideoxy-*D*-glycero- α -*D*-galacto-2-nonulopyranosid)onate (17a). Compound **16a** (55.0 mg, 0.16 mmol) was

dissolved in dry acetonitrile (2 mL). Powdered MS 3 Å (50 mg) and 2,3-difluorobenzyl alcohol (35.0 μ L, 42.0 mg, 0.29 mmol) were added. The reaction mixture was stirred at rt for 1.5 h. Then the suspension was cooled to -40 °C and was subsequently treated with *N*-iodosuccinimide (35.0 mg, 0.16 mmol) and triflic acid (8.00 μ L, 13.0 mg, 0.09 mmol). After 30 min, the reaction mixture was warmed to -30 °C and stirring was continued for 20 h. After warming to rt, the mixture was filtered through a pad of celite and washed with 20% $\text{Na}_2\text{S}_2\text{O}_3$ (1 \times 2 mL), saturated aqueous NaHCO_3 (3 \times 5 mL), and H_2O (1 \times 5 mL). The organic phase was dried over Na_2SO_4 , filtered, and concentrated under reduced pressure. The crude product was purified by chromatography on silica gel (1% gradient *i*PrOH in petrol ether/DCM 2:1) to yield **16a** (43 mg, 66%) as a colorless oil. $[\alpha]_{\text{D}}^{20}$ -0.04 (c 0.73, CH_2Cl_2). ^1H NMR (500 MHz, CDCl_3) δ 2.03 (s, 3H, OAc), 2.04 (m, 1H, H-3a), 2.16, 2.21 (2 s, 6H, 2 OAc), 2.69 (dd, J = 4.7, 12.9 Hz, 1H, H-3b), 3.26 (dd, J = 5.3, 13.5 Hz, 1H, H-9a), 3.58 (dd, J = 2.8, 13.5 Hz, 1H, H-9b), 3.79 (s, 3H, OMe), 4.21 (m, 2H, H-5, H-6), 4.53 (A of AB, J = 12.0 Hz, 1H, CH_2Ar), 4.63 (m, 2H, CH_2F), 4.83 (B of AB, J = 12.3 Hz, 1H, CH_2Ar), 4.95 (ddd, J = 4.7, 10.0, 12.1 Hz, 1H, H-4), 5.35 (m, 2H, H-7, H-8), 6.18 (dd, J = 3.3, 9.0 Hz, 1H, *NHFAc*), 7.11 (m, 3H, CH_{ar}). ^{13}C NMR (CDCl_3) δ 20.9 (3C, OAc), 37.9 (C-3), 48.6 (C-5), 50.9 (C-9), 53.0 (OMe), 60.5 (CH_2Ar), 67.8 (C-7), 68.7 (C-8), 69.3 (C-4), 72.6 (C-6), 80.1 (J = 186.1 Hz, CH_2F), 98.7 (C-2), 116.9, 117.1, 125.1, 126.6 (6C, C-Ar), 168.0, 170.2, 170.7 (5C, CO). ESI-MS calcd for $\text{C}_{25}\text{H}_{29}\text{F}_3\text{N}_4\text{O}_{11}[\text{M} + \text{Na}]^+$ 641.17; found m/z 641.14.

Methyl ((2,3-Difluorobenzyl) 5-Chloroacetamido-4,7,8-tri-*O*-acetyl-9-azido-3,5,9-trideoxy-*D*-glycero- α -*D*-galacto-2-nonulopyranosid)onate (17b). Compound **16b** (63.0 mg, 0.12 mmol) was reacted with 2,3-difluorobenzyl alcohol (41.0 μ L, 0.36 mmol), *N*-iodosuccinimide (42.0 mg, 0.19 mmol), and triflic acid (8.0 μ L, 14.4 mg, 0.1 mmol). The crude product was purified by chromatography on silica gel (0.5% gradient *i*PrOH in PE:DCM, 8:4) to yield **17b** (50 mg, 68%) as a colorless oil. $[\alpha]_{\text{D}}^{20}$ -0.01 (c 1.05, CH_2Cl_2). ^1H NMR (500 MHz, CDCl_3) δ 2.02 (m, 1H, H-3a), 2.03, 2.16, 2.21 (3 s, 9H, 3 OAc), 2.70 (dd, J = 4.6, 12.9 Hz, 1H, H-3b), 3.28 (dd, J = 5.8, 13.5 Hz, 1H, H-9a), 3.59 (dd, J = 3.0, 13.5 Hz, 1H, H-9b), 3.81 (s, 3H, OMe), 3.93, 4.01 (A, B of AB, J = 15.0 Hz, 2H, CH_2Cl), 4.11 (m, 1H, H-5), 4.24 (dd, J = 2.1, 10.7 Hz, 1H, H-6), 4.54, 4.83 (A, B of AB, J = 12.0 Hz, 2H, CH_2Ar), 4.99 (m, 1H, H-4), 5.32 (dd, J = 2.1, 7.8 Hz, 1H, H-7), 5.36 (m, 1H, H-8), 6.41 (d, J = 10.1 Hz, 1H, NH), 7.06-7.17 (m, 3H, CH_{ar}). ^{13}C NMR (CDCl_3) δ 20.8, 20.9, 21.1 (3 OAc), 37.9 (C-3), 42.4 (CH_2Cl), 49.7 (C-5), 50.9 (C-9), 60.5 (CH_2Ar), 67.9, 68.1, 68.4 (3C, C-4, C-7, C-8), 72.6 (C-6), 98.7 (C-2), 117.0, 123.9, 125.1, 126.5, 149.5, 151.4 (6 C-Ar), 166.7, 168.0, 170.2, 170.3, 170.7 (5 CO). ESI-MS calcd for $\text{C}_{25}\text{H}_{29}\text{Cl}_2\text{FN}_4\text{O}_{11}[\text{M} + \text{Na}]^+$ 657.14; found m/z 657.29.

Methyl ((2,3-Difluorobenzyl) 5-(*o*-Nitrotoluenesulfonamido)-4,7,8-tri-*O*-acetyl-9-azido-3,5,9-trideoxy-*D*-glycero- α -*D*-galacto-2-nonulopyranosid)onate (17c). Compound **16c** (62.0 mg, 0.10 mmol) was reacted with 2,3-difluorobenzyl alcohol (30.0 μ L, 38.0 mg, 0.29 mmol), *N*-iodosuccinimide (32.0 mg, 0.14 mmol), and triflic acid (7.00 μ L, 12.0 mg, 0.08 mmol). The crude product was purified by chromatography on silica gel (toluene/EA, gradient 1:1 to 2:1) to yield **17c** (44 mg, 61%). ^1H NMR (500 MHz, CDCl_3) δ 1.84 (t, J = 12.4 Hz, 1H, H-3a), 2.08, 2.17, 2.28 (3 s, 9H, 3 OAc), 2.64 (dd, J = 4.6, 12.8 Hz, 1H, H-3b), 3.29 (m, 1H, H-9a), 3.45 (m, 1H, H-9b), 3.73 (s, 3H, OMe), 3.79 (m, 2H, H-5, H-7), 4.12 (d, J = 8.8 Hz, 1H, H-6), 4.46, 4.77 (A, B of AB, J = 12.0 Hz, 2H, CH_2Ar), 4.91 (td, J = 4.6, 11.5 Hz, 1H, H-4), 5.25 (d, J = 7.8 Hz, 1H, H-8), 5.60 (d, J = 9.4 Hz, 1H, NH), 7.09 (m, 3H, CH_{ar}), 7.69 (m, 2H, CH_{ar}), 7.85 (d, J = 7.9 Hz, 1H, CH_{ar}), 8.07 (d, J = 7.8 Hz, 1H, CH_{ar}). ^{13}C NMR (CDCl_3) δ 21.1 (3C, OAc), 38.1 (C-3), 51.1 (C-5), 53.1 (C-9), 53.6 (OMe), 60.5 (CH_2Ar), 68.7, 68.8 (2C, C-4, C-7), 69.8 (C-8), 73.2 (C-6), 125.5, 130.4, 133.5, 147.6 (12C, C-Ar), 168.0, 170.1, 170.3, 170.5 (4 CO). ESI-MS calcd for $\text{C}_{29}\text{H}_{31}\text{F}_2\text{N}_5\text{O}_{14}\text{S}[\text{M} + \text{Na}]^+$ 766.16; found m/z 766.14.

Methyl ((2,3-Difluorobenzyl) 5-Methylsulfonamido-4,7,8-tri-O-acetyl-9-azido-3,5,9-trideoxy-D-glycero- α -D-galacto-2-nonulopyranosid)onate (17d). Compound **16d** (56.0 mg, 0.10 mmol) was reacted with 2,3-difluorobenzyl alcohol (33.0 μ L, 42.0 mg, 0.29 mmol), *N*-iodosuccinimide (35.0 mg, 0.16 mmol), and triflic acid (8.00 μ L, 13.0 mg, 0.09 mmol). The crude product was purified by chromatography on silica gel (1% gradient *i*PrOH in petrol ether/DCM 2:1) to yield **17d** (37 mg, 56%). ¹H NMR (500 MHz, CDCl₃) δ 1.95 (t, *J* = 12.5, 1H, H-3a), 2.13, 2.18 (2 s, 9H, 3 OAc), 2.61 (dd, *J* = 4.6, 12.7 Hz, 1H, H-3b), 3.02 (s, 3H, CH₃), 3.31 (dd, *J* = 6.3, 13.5 Hz, 1H, H-9a), 3.60 (dd, *J* = 3.3, 13.5 Hz, 1H, H-9b), 3.71 (m, 4H, H-5, OMe), 4.12 (dd, *J* = 1.7, 10.7 Hz, 1H, H-6), 4.54, 4.81 (A, B of AB, *J* = 11.8 Hz, 2H, CH₂Ar), 5.09 (m, 1H, H-4), 5.31 (m, 1H, H-8), 5.49 (dd, *J* = 1.7, 7.5 Hz, 1H, H-7), 7.11 (m, 3H, CH_{ar}). ¹³C NMR (CDCl₃) δ 20.9, 21.1, 21.4 (3 OAc), 38.1 (C-3), 42.3 (CH₃), 50.8 (C-9), 52.6, 53.0 (2C, OMe, C-5), 60.4 (CH₂Ar), 68.1 (C-4), 68.8 (C-7), 70.0 (C-8), 72.8 (C-6), 98.5 (C-2), 116.9, 117.0, 124.0, 125.1 (6C, C-Ar), 167.7, 170.1, 170.7, 172.1 (4 CO). ESI-MS calcd for C₂₄H₃₀F₂N₄O₁₂S [M + Na]⁺ 659.15; found *m/z* 659.24.

Methyl ((2,3-Difluorobenzyl) 5-Fluoroacetamido-4,7,8-tri-O-acetyl-9-(4-chlorobenzamido)-3,5,9-trideoxy-D-glycero- α -D-galacto-2-nonulopyranosid)onate (18a). Compound **17a** (55.0 mg, 0.09 mmol) was dissolved in dry DCM (2 mL). *p*-Chlorobenzoyl chloride (45.0 μ L, 62.0 mg, 0.36 mmol) and triphenylphosphine (47.0 mg, 0.18 mmol) were added. The reaction mixture was stirred at rt overnight. Afterward, the reaction mixture was washed with saturated aqueous NaHCO₃ (3 \times 5 mL) and H₂O (1 \times 5 mL). The organic phase was dried over Na₂SO₄, filtered, and concentrated under reduced pressure. The crude product was purified by chromatography on silica gel (EA:PE, gradient 1:1 to 2:1) to yield **18a** (31 mg, 48%) as a white foam. ¹H NMR (500 MHz, CDCl₃) δ 1.97 (m, 1H, H-3a), 1.97, 2.09, 2.18 (3 s, 9H, 3 OAc), 2.64 (dd, *J* = 4.6, 12.8 Hz, 1H, H-3b), 2.94 (dd, *J* = 3.9, 15.0 Hz, 1H, H-9a), 3.69 (s, 3H, OMe), 4.09 (m, 1H, H-6), 4.22 (m, 2H, H-5, H-9b), 4.46 (A of AB, *J* = 12.0 Hz, 1H, CH₂Ar), 4.64 (m, 2H, CH₂F), 4.78 (B of AB, *J* = 12.0 Hz, 1H, CH₂Ar), 4.86 (m, 1H, H-4), 5.11 (dd, *J* = 2.0, 9.8 Hz, 1H, H-7), 5.27 (m, 1H, H-8), 6.12 (dd, *J* = 3.3, 10.0 Hz, 1H, NHFAc), 6.89 (dd, *J* = 4.3, 7.9 Hz, 1H, NH), 7.03 (m, 3H, CH_{ar}), 7.34, 7.67 (AA', BB' of AA'/BB', *J* = 8.5 Hz, 4H, CH_{ar}). ¹³C NMR (CDCl₃) δ 20.8, 21.1 (3C, OAc), 37.9 (C-3), 38.7 (C-9), 48.7 (C-5), 52.9 (OMe), 60.5 (CH₂Ar), 67.8 (C-7), 68.2 (C-8), 68.7 (C-4), 72.0 (C-6), 80.0 (*J* = 186.0 Hz, CH₂F), 98.6 (C-2), 117.1, 124.0, 125.3, 126.6, 128.4, 128.8, 131.5, 132.6, 137.8 (12C, C-Ar), 166.4, 167.8, 168.3, 170.5, 170.7, 172.3 (6 CO). ESI-MS calcd for C₃₂H₃₄ClF₃N₂O₁₂ [M + Na]⁺ 753.18; found *m/z* 753.19.

Methyl ((2,3-Difluorobenzyl) 5-Chloroacetamido-4,7,8-tri-O-acetyl-9-(4-chloro-benzamido)-3,5,9-trideoxy-D-glycero- α -D-galacto-2-nonulopyranosid)onate (18b). Compound **17b** (52.2 mg, 0.08 mmol) was reacted with *p*-chlorobenzoyl chloride (23.0 μ L, 31.0 mg, 0.18 mmol) and triphenylphosphine (47.0 mg, 0.18 mmol). The crude product was purified by chromatography on silica gel (0.5% gradient of MeOH in DCM) to yield in **18b** (25.0 mg, 48%). [α]_D²⁰ 0.04 (*c* 1.06, CH₂Cl₂). ¹H NMR (500 MHz, CDCl₃) δ 2.00 (m, 1H, H-3a), 2.03, 2.16, 2.25 (3 s, 9H, 3 OAc), 2.71 (m, 1H, H-3b), 3.01 (m, 1H, H-9a), 3.77 (s, 3H, OMe), 3.97 (m, 2H, CH₂Cl), 4.11 (m, 1H, H-6), 4.20 (m, 1H, H-5), 4.29 (m, 1H, H-9b), 4.54, 4.85 (A, B of AB, *J* = 11.9 Hz, 2H, CH₂Ar), 4.95 (m, 1H, H-4), 5.18 (m, 1H, H-7), 5.31 (m, 1H, H-8), 6.39 (m, 1H, NHAc), 6.97 (m, 1H, NH), 7.06–7.15 (m, 3H, CH_{ar}), 7.40 (m, 2H, CH_{ar}), 7.74 (m, 2H, CH_{ar}). ¹³C NMR (CDCl₃) δ 20.8, 21.2 (3C, OAc), 38.0 (C-3), 38.6 (C-9), 42.4 (CH₂Cl), 49.8 (C-5), 52.9 (OMe), 60.4 (CH₂Ar), 67.8, 68.3, 68.5 (3C, C-4, C-7, C-8), 72.1 (C-6), 98.6 (C-2), 117.0 (*J* = 17.5 Hz), 124.0, 125.3, 126.5, 128.4, 128.9, 132.7, 137.8 (12C-Ar), 166.4, 166.7, 167.9, 170.5, 170.7, 172.3 (6 CO). ESI-MS calcd for C₃₂H₃₄Cl₂F₂N₂O₁₂ [M + Na]⁺ 769.14; found *m/z* 769.34.

Methyl ((2,3-Difluorobenzyl) 5-(*o*-Nitrotoluenesulfonamido)-4,7,8-tri-O-acetyl-9-(4-chlorobenzamido)-3,5,9-trideoxy-D-glycero- α -D-galacto-2-nonulopyranosid)onate (18c). Compound

17c (44.0 mg, 0.06 mmol) was dissolved in dry DCE (2 mL). *p*-Chlorobenzoyl chloride (30.0 μ L, 40.0 mg, 0.23 mmol) and triphenylphosphine (31.0 mg, 0.12 mmol) were added. The reaction mixture was stirred at rt overnight. The crude product was purified by chromatography on silica gel (EA:PE, gradient 1:1 to 2:1) to yield **18c** (29 mg, 58%) as a white foam. ¹H NMR (500 MHz, CDCl₃) δ 1.90 (m, 1H, H-3a), 2.04, 2.14, 2.32 (3 s, 9H, 3 OAc), 2.69 (dd, *J* = 4.5, 12.7 Hz, 1H, H-3b), 2.92 (d, *J* = 15.2 Hz, 1H, H-9a), 3.75 (s, 3H, OMe), 3.91 (m, 1H, H-5), 4.12 (m, 1H, H-6), 4.38 (m, 1H, H-9b), 4.52 (A of AB, *J* = 11.9 Hz, 1H, CH₂Ar), 4.84 (m, 2H, H-4, CH₂Ar), 5.31 (m, 2H, H-7, H-8), 5.64 (d, *J* = 9.4 Hz, 1H, NH), 7.13 (m, 4H, CH_{ar}), 7.21 (m, 1H, NH), 7.40 (m, 3H, CH_{ar}), 7.72 (m, 4H, CH_{ar}). ¹³C NMR (CDCl₃) δ 20.3, 21.2, 21.3 (3 OAc), 38.2, 38.3 (2C, C-3, C-9), 53.0 (OMe), 53.7 (C-5), 60.6 (CH₂Ar), 68.2, 68.4 (2C, C-7, C-8), 69.0 (C-4), 72.6 (C-6), 98.2 (C-2), 117.0, 123.9, 125.6, 128.8, 130.3, 131.5, 132.1, 132.8, 133.3 (18C, C-Ar), 166.6, 167.7, 169.9, 170.4, 172.4 (5 CO). ESI-MS calcd for C₃₆H₃₆ClF₂N₃O₁₅S [M + Na]⁺ 878.15; found *m/z* 878.28.

Methyl ((2,3-Difluorobenzyl) 5-Methylsulfonamido-4,7,8-tri-O-acetyl-9-(4-chlorobenzamido)-3,5,9-trideoxy-D-glycero- α -D-galacto-2-nonulopyranosid)onate (18d). Compound **17d** (35.0 mg, 0.06 mmol) was dissolved in dry DCM (2 mL). *p*-Chlorobenzoyl chloride (28.0 μ L, 38.0 mg, 0.22 mmol) and triphenylphosphine (29.0 mg, 0.11 mmol) were added. The reaction mixture was stirred at rt overnight. The crude product was purified by chromatography on silica gel (EA:PE, gradient 1:1 to 2:1) to yield **18d** (21 mg, 52%) as a white foam. ¹H NMR (500 MHz, CDCl₃) δ 1.98 (t, *J* = 13.6 Hz, 1H, H-3a), 2.05, 2.10, 2.13 (3 s, 9H, 3 OAc), 2.67 (dd, *J* = 4.6, 12.7 Hz, 1H, H-3b), 2.99 (s, 3H, CH₃), 3.05 (m, 1H, H-9a), 3.70 (s, 3H, OMe), 3.77 (m, 1H, H-5), 4.09 (d, *J* = 1.5 Hz, 1H, H-6), 4.30 (ddd, *J* = 3.0, 8.2, 15.1 Hz, 1H, H-9b), 4.54 (A of AB, *J* = 12.1 Hz, 1H, CH₂Ar), 4.59 (d, *J* = 9.6 Hz, 1H, NH), 4.83 (B of AB, *J* = 11.9 Hz, 1H, CH₂Ar), 4.99 (ddd, *J* = 4.6, 10.4, 12.2 Hz, 1H, H-4), 5.30 (m, 1H, H-8), 5.37 (dd, *J* = 1.7, 9.6 Hz, 1H, H-7), 7.13 (m, 4H, CH_{ar}, NH), 7.40, 7.75 (AA', BB' of AA'/BB', *J* = 8.5 Hz, 4H, CH_{ar}). ¹³C NMR (CDCl₃) δ 21.2, 21.2, 21.3 (3 OAc), 29.7 (C-3), 38.2, 38.8 (2C, C-9, CH₃), 42.3 (C-5), 52.8 (OMe), 60.5 (CH₂Ar), 68.1, 68.5, 69.1 (3C, C-4, C-7, C-8), 72.4 (C-6), 98.4 (C-2), 116.9, 117.1, 125.3, 128.4, 128.9, 132.1, 132.7, 137.8 (12C, C-Ar), 166.4, 167.7, 170.5, 171.6, 172.6 (5 CO). ESI-MS calcd for C₃₁H₃₅ClF₂N₂O₁₃S [M + Na]⁺ 771.15; found *m/z* 771.29.

Sodium ((2,3-Difluorobenzyl) 5-Fluoroacetamido-9-(4-chlorobenzamido)-3,5,9-trideoxy-D-glycero- α -D-galacto-2-nonulopyranosid)onate (19a). Compound **18a** (28.0 mg, 0.04 mmol) was dissolved in THF/water (2.0 mL/0.5 mL) and was reacted with LiOH (9.00 mg, 0.38 mmol). The reaction mixture was stirred at rt for 4 h. 7% HCl (aq) was added to adjust the pH to 7. The crude product was purified by reversed-phase chromatography (RP-18, 10% gradient MeOH in H₂O) followed by ion exchange chromatography (Dowex-50) and P2 size exclusion chromatography to yield **19a** (7.0 mg, 30%) as a white foam. [α]_D²⁰ -0.73 (*c* 0.08, H₂O). ¹H NMR (500 MHz, D₂O) δ 1.59 (t, *J* = 12.2, 1H, H-3a), 2.65 (dd, *J* = 4.7, 12.4, 1H, H-3b), 3.31 (dd, *J* = 7.7, 14.0 Hz, 1H, H-9a), 3.41 (dd, *J* = 1.8, 8.8 Hz, 1H, H-5), 3.64 (m, 3H, H-4, H-8, H-9b), 3.77 (dd, *J* = 1.9, 10.5 Hz, 1H, H-6), 3.84 (m, 1H, H-7), 4.52 (A of AB, *J* = 11.7 Hz, 1H, CH₂Ar), 4.70 (m, 2H, CH₂Ar, CH₂F), 4.81 (d, *J* = 5.1 Hz, 1H, CH₂F), 7.02 (m, 3H, CH_{ar}), 7.39, 7.62 (AA', BB' of AA'/BB', *J* = 8.7 Hz, 4H, CH_{ar}). ¹³C NMR (CDCl₃) δ 40.4, 42.6 (2C, C-3, C-9), 57.2 (C-5), 60.6 (CH₂Ar), 68.0 (C-4), 69.7 (C-7), 70.3 (C-8), 72.3 (C-6), 87.0 (CH₂F), 101.3 (C-2), 117.2, 125.8, 128.6, 128.8, 132.1, 137.5, 141.7 (12C, C-Ar), 173.1, 175.3, 189.7 (3 CO). HRMS calcd for C₂₅H₂₆ClF₃N₂O₉ [M - H]⁻ 589.1206; found 589.1191.

Sodium ((2,3-Difluorobenzyl) 5-Chloroacetamido-9-(4-chlorobenzamido)-3,5,9-trideoxy-D-glycero- α -D-galacto-2-nonulopyranosid)onate (19b). Compound **18b** (21.0 mg, 0.03 mmol) was treated with LiOH (6.7 mg, 0.3 mmol) in THF/water (2.0 mL/0.5 mL). The crude product was purified by chromatography on silica gel

(0.1% gradient of H₂O in DCM/MeOH; 2:1) followed by ion exchange chromatography (Dowex 50) and P2 size exclusion chromatography to yield **19b** as a white solid (10 mg, 60%). ¹H NMR (500 MHz, MeOD) δ 1.58 (dd, *J* = 10.7, 12.1 Hz, 1H, H-3a), 2.82 (m, 1H, H-3b), 3.35 (m, 1H, H-7), 3.46 (m, 1H, H-9a), 3.68 (m, 4H, H-4, H-5, H-6, H-9b), 3.93 (m, 2H, CH₂Cl), 4.57, 4.84 (A, B of AB, *J* = 12.1 Hz, 2H, CH₂Ar), 7.01 (m, 3H, CH_{ar}), 7.35, 7.72 (AA', BB' of AA'/BB', *J* = 8.5 Hz, 4H, CH_{ar}). ¹³C NMR (MeOD) δ 43.0 (C-3), 44.8 (C-9), 45.0 (CH₂Cl), 54.0 (C-5), 63.1 (CH₂Ar), 70.5, 72.2, 72.9 (3C, C-4, C-7, C-8), 74.9 (C-6), 103.8 (C-2), 119.7, 126.9, 128.4, 128.9, 131.2, 131.3, 140.1 (10 C-Ar), 170.0, 170.5, 173.11 (3 CO). HRMS calcd for C₂₅H₂₅Cl₂F₂N₂NaO₉ [M + Na]⁺ 651.0701; found *m/z* 651.0700.

Sodium ((2,3-Difluorobenzyl) 5-(*o*-Nitrotoluenesulfonamido)-9-(4-chlorobenzamido)-3,5,9-trideoxy-*D*-glycero- α -*D*-galacto-2-nonulopyranosid)onate (19c). Compound **18c** (29.0 mg, 0.03 mmol) was dissolved in THF/H₂O (2 mL/0.5 mL) and was reacted with LiOH (8.00 mg, 0.33 mmol). The crude product was purified on RP-18 (10% gradient MeOH in water) followed by ion exchange chromatography (Dowex-50) and P2 size exclusion chromatography to yield **19c** (10 mg, 40%). [α]_D²⁰ -0.35 (c 0.10, H₂O). ¹H NMR (500 MHz, D₂O) δ 1.45 (t, *J* = 12.2 Hz, 1H, H-3a), 2.52 (dd, *J* = 4.5, 12.3 Hz, 1H, H-3b), 3.16 (t, *J* = 9.8 Hz, 1H, H-5), 3.43 (m, 2H, H-4, H-9a), 3.60 (dd, *J* = 2.9, 14.2 Hz, 1H, H-9b), 3.67 (m, 2H, H-6, H-8), 3.74 (dd, *J* = 1.1, 8.9 Hz, 1H, H-7), 4.50 (A of AB, *J* = 11.7 Hz, 1H, CH₂Ar), 4.65 (m, 1H, CH₂Ar), 7.04 (m, 3H, CH_{ar}), 7.42 (AA' of AA'/BB', *J* = 8.5 Hz, 2H, CH_{ar}), 7.60 (m, 4H, CH_{ar}), 7.98 (BB' of AA'/BB', *J* = 7.6 Hz, 2H, CH_{ar}). ¹³C NMR (D₂O) δ 40.4 (C-3), 42.4 (C-9), 56.5 (C-5), 60.5 (CH₂Ar), 69.4 (C-7), 69.5 (C-4), 70.3 (C-8), 73.8 (C-6), 101.2 (C-2), 117.1, 124.2, 125.8, 128.7, 130.2, 132.1, 132.5, 137.5, 146.8 (18C, C-Ar), 170.0, 173.3 (2 CO). HRMS calcd for C₂₉H₂₇ClF₂N₃NaO₁₂S [M + Na]⁺ 760.0767; found *m/z* 760.0775.

Sodium ((2,3-Difluorobenzyl) 5-Methylsulfonamido-9-(4-chlorobenzamido)-3,5,9-trideoxy-*D*-glycero- α -*D*-galacto-2-nonulopyranosid)onate (19d). Compound **18d** (21.0 mg, 0.03 mmol) was dissolved in THF/H₂O (2 mL/0.5 mL) and was reacted with LiOH (7.00 mg, 0.28 mmol). The crude product was purified on RP-18 (10% gradient MeOH in water) followed by ion exchange chromatography (Dowex-50) and P2 size exclusion chromatography to yield **19d** (10 mg, 59%). [α]_D²⁰ -0.44 (c 0.07, H₂O). ¹H NMR (500 MHz, D₂O) δ 1.56 (t, *J* = 12.2 Hz, 1H, H-3a), 2.62 (dd, *J* = 4.6, 12.4 Hz, 1H, H-3b), 3.02 (s, 3H, CH₃), 3.23 (t, *J* = 10.1 Hz, 1H, H-5), 3.40 (dd, *J* = 7.8, 14.7 Hz, 1H, H-9a), 3.48 (td, *J* = 4.6, 11.9 Hz, 1H, H-4), 3.66 (m, 3H, H-6, H-8, H-9b), 3.75 (d, *J* = 8.7 Hz, 1H, H-7), 4.51 (A of AB, *J* = 11.8 Hz, 1H, CH₂Ar), 4.66 (m, 1H, CH₂Ar), 7.02 (m, 3H, CH_{ar}), 7.40, 7.63 (AA', BB' of AA'/BB', *J* = 8.6 Hz, 4H, CH_{ar}). ¹³C NMR (D₂O) δ 40.9, 41.3 (2C, C-3, C-9), 42.5 (CH₃), 55.5 (C-5), 60.6 (CH₂Ar), 68.6 (C-7), 69.3 (C-4), 70.3 (C-8), 73.2 (C-6), 101.2 (C-2), 112.4, 117.2, 124.4, 125.8, 128.7, 128.8, 137.5 (12C, C-Ar), 173.1, 187.7 (2 CO). HRMS calcd for C₂₄H₂₆ClF₂N₂NaO₁₀S [M + Na]⁺ 653.0760; found *m/z* 653.0759.

Sodium ((2,3-Difluorobenzyl) 5-Methoxyacetamido-9-(4-chlorobenzamido)-3,5,9-trideoxy-*D*-glycero- α -*D*-galacto-2-nonulopyranosid)onate (19e). Compound **18b** (10 mg, 0.01 mmol) was treated with 10% aqueous NaOH (0.3 mL) in MeOH (1.5 mL) at rt for 4 h. The crude product was purified by chromatography on RP-18 (5% gradient of MeOH in H₂O) to yield **19e** (2.3 mg, 19%) and **19b** (1.7 mg, 21%) as white solids. [α]_D²⁰ -19.1 (c 0.46, MeOH). ¹H NMR (500 MHz, D₂O) δ 1.67 (t, *J* = 12.2 Hz, 1H, H-3a), 2.73 (dd, *J* = 4.6, 12.4 Hz, 1H, H-3b), 3.28 (s, 3H, OMe), 3.41-3.56 (m, 2H, H-7, H-9a), 3.62-3.79 (m, 3H, H-4, H-8, H-9b), 3.82-3.89 (m, 2H, H-5, H-6), 3.95, 3.98 (A, B of AB, *J* = 15.6 Hz, 2H, CH₂OMe), 4.62, 4.78 (A, B of AB, *J* = 11.7 Hz, 2H, CH₂Ar), 7.00-7.22 (m, 3H, CH_{ar}), 7.49, 7.71 (AA', BB' of AA'/BB', *J* = 8.5 Hz, 4H, CH_{ar}). ¹³C NMR (D₂O) δ 40.4 (C-3), 42.4 (C-9), 51.5 (C-5), 58.9 (OMe), 60.6 (CH₂Ar), 67.9 (C-4), 69.6 (C-7), 70.2 (C-8), 70.8 (CH₂OMe), 72.4 (C-6), 101.2 (C-2), 110.0, 117.1, 117.2, 124.4, 125.9, 128.7, 128.8, 132.1 (12C, C-Ar), 173.0,

173.5 (3C, CO). HRMS calcd for C₂₆H₂₈ClF₂N₂O₁₀Na [M + Na]⁺ 647.1195; found *m/z* 647.5573.

Sodium ((2,3-Difluorobenzyl) 5-Cyclopropylamido-9-(4-chlorobenzamido)-3,5,9-trideoxy-*D*-glycero- α -*D*-galacto-2-nonulopyranosid)onate (19f). Compound **23f** (31.0 mg, 0.04 mmol) was dissolved in THF/water (2.0 mL/0.5 mL) and was reacted with LiOH (10.0 mg, 0.42 mmol). The crude product was purified on RP-18 (10% gradient MeOH in H₂O) followed by ion exchange chromatography (Dowex-50) and P2 size exclusion chromatography to yield **19f** (11 mg, 44%). [α]_D²⁰ -0.1 (c 0.26, H₂O). ¹H NMR (500 MHz, D₂O) δ 0.59 (m, 2H, CH₂), 0.70 (m, 2H, CH₂), 1.44 (m, 1H, CH), 1.56 (t, *J* = 11.8 Hz, 1H, H-3a), 2.64 (dd, *J* = 3.5, 11.8 Hz, 1H, H-3b), 3.37 (d, *J* = 8.2 Hz, 1H, H-7), 3.47 (m, 1H, H-9a), 3.63 (m, 4H, H-5, H-6, H-8, H-9b), 4.52, 4.69 (A, B of AB, *J* = 11.6 Hz, 2H, CH₂Ar), 7.04 (m, 3H, CH_{ar}), 7.40, 7.63 (AA', BB' of AA'/BB', *J* = 8.1 Hz, 4H, CH_{ar}). ¹³C NMR (CDCl₃) δ 6.8, 7.0 (2C, CH₂), 14.1 (CH), 40.5, 42.3 (2C, C-3, C-9), 51.9 (C-5), 60.5 (CH₂Ar), 68.0 (C-4), 69.4 (C-7), 70.1 (C-8), 72.8 (C-6), 101.2 (C-2), 117.1, 124.4, 125.9, 128.7, 132.1, 137.6 (12C, C-Ar), 170.0, 173.1, 178.3 (3 CO). HRMS calcd for C₂₇H₂₈ClF₂N₂O₉ [M - H]⁻ 597.1457; found 597.1454.

Sodium ((2,3-Difluorobenzyl) 5-Cyclobutylamido-9-(4-chlorobenzamido)-3,5,9-trideoxy-*D*-glycero- α -*D*-galacto-2-nonulopyranosid)onate (19g). Compound **23g** (44.0 mg, 0.06 mmol) was dissolved in THF/water (2.0 mL/0.5 mL) and was reacted with LiOH (14.0 mg, 0.58 mmol). The crude product was purified on RP-18 (10% gradient MeOH in H₂O) followed by ion exchange chromatography (Dowex-50) and P2 size exclusion chromatography to yield **19g** (17 mg, 49%). [α]_D²⁰ -0.2 (c 0.11, H₂O). ¹H NMR (500 MHz, D₂O) δ 1.55 (m, 1H, H-3a), 1.85 (m, 6H, CH₂), 2.63 (dd, *J* = 3.9, 12.2 Hz, 1H, H-3b), 2.99 (quint, *J* = 8.8 Hz, 1H, CH), 3.32 (d, *J* = 8.6 Hz, 1H, H-9a), 3.61 (m, 6H, H-4, H-5, H-6, H-7, H-8, H-9b), 4.52, 4.68 (A, B of AB, *J* = 11.4 Hz, 2H, CH₂Ar), 7.05 (m, 3H, CH_{ar}), 7.40, 7.63 (AA', BB' of AA'/BB', *J* = 7.5 Hz, 4H, CH_{ar}). ¹³C NMR (D₂O) δ 17.4, 24.5 (3C, CH₂), 39.2 (CH), 40.5 (C-3), 42.1 (C-9), 51.6 (C-5), 60.5 (CH₂Ar), 67.8 (C-4), 69.3 (C-7), 70.1 (C-8), 72.7 (C-6), 110.0 (C-2), 115.6, 123.3, 125.9, 128.8, 150.9 (12C, C-Ar), 170.0, 175.8, 179.7 (3 CO). ESI-MS calcd for C₂₈H₃₀ClF₂N₂NaO₉ [M + Na]⁺ 657.1403; found *m/z* 657.1401.

Methyl (Methyl 5-*tert*-Butyloxycarbonylamino-4,7,8-tri-*O*-acetyl-9-(4-chlorobenzamido)-3,5,9-trideoxy-2-thio-*D*-galacto- α -*D*-galacto-2-nonulopyranosid)onate (20). Compound **14** (167 mg, 0.30 mmol) was dissolved in dry DCE (5 mL). *p*-Chlorobenzoyl chloride (150 μ L, 210 mg, 1.19 mmol) and triphenylphosphine (156 mg, 0.59 mmol) were added. The reaction mixture was stirred at rt overnight. Afterward, the reaction mixture was washed with saturated aqueous NaHCO₃ (3 \times 5 mL) and H₂O (1 \times 5 mL). The organic phase was dried over Na₂SO₄, filtered, and concentrated under reduced pressure. The crude product was purified by chromatography on silica gel (EA:PE, gradient 1:1 to 2:1) to yield **20** (95 mg, 48%) as a white foam. ¹H NMR (500 MHz, CD₂Cl₂) δ 1.39 (s, 9H, C(CH₃)₃), 1.92 (s, 3H, SME), 2.01 (s, 6H, 2 OAc), 2.12 (m, 1H, H-3a), 2.23 (s, 3H, OAc), 2.56 (dd, *J* = 4.7, 13.8 Hz, 1H, H-3b), 3.00 (td, *J* = 3.2, 15.2 Hz, 1H, H-9a), 3.79 (m, 4H, OMe, H-5), 4.23 (d, *J* = 10.6 Hz, 1H, H-6), 4.41 (ddd, *J* = 3.1, 8.8, 14.4 Hz, 1H, H-9b), 4.52 (dd, *J* = 2.9, 10.0 Hz, 1H, NHAc), 5.09 (m, 1H, H-8), 5.16 (td, *J* = 4.6, 11.0 Hz, 1H, H-4), 5.33 (m, 1H, H-7), 7.23 (d, *J* = 3.9 Hz, 1H, NH), 7.54, 8.08 (AA', BB' of AA'/BB', *J* = 8.5 Hz, 4H, CH_{ar}). ¹³C NMR (CD₂Cl₂) δ 11.7 (SME), 21.2, 21.3, 21.4 (3 OAc), 28.5 (C(CH₃)₃), 37.8 (C-3), 38.4 (C-9), 51.5 (C-5), 53.3 (OMe), 68.9 (C-7), 70.0 (C-4), 70.6 (C-8), 71.5 (C-6), 80.5 (C-2), 85.3 (C(CH₃)₃), 129.3, 132.4, 138.0, 141.8 (6C, C-Ar), 155.7 (CONH), 166.7, 168.6, 170.5, 170.7, 172.4 (5 CO). ESI-MS calcd for C₂₉H₃₀ClN₂O₁₂S [M + Na]⁺ 697.18; found *m/z* 697.25.

Methyl (Methyl 5-Amino-4,7,8-tri-*O*-acetyl-9-(4-chlorobenzamido)-3,5,9-trideoxy-2-thio-*D*-glycero- α -*D*-galacto-2-nonulopyranosid)onate (21). Compound **20** (95.0 mg, 0.14 mmol) was dissolved in 4 M PhOH (in DCM; 4 mL) and 4 M TMSCl

(in DCM; 2 mL). The reaction mixture was stirred at rt for 2 h. The crude product was purified by chromatography on silica gel (EA:PE, gradient 1:1 to 8:1) to yield **21** as a white foam (51 mg, 63%). ¹H NMR (500 MHz, CD₂Cl₂) δ 1.95 (s, 3H, SMe), 1.99 (dd, *J* = 5.7, 14.0 Hz, 1H, H-3a), 2.04, 2.06, 2.19 (3 s, 9H, 3 OAc), 2.57 (dd, *J* = 4.7, 13.7 Hz, 1H, H-3b), 2.62 (t, *J* = 10.0 Hz, 1H, H-5), 3.34–3.42 (m, 1H, H-9a), 3.76 (s, 3H, OMe), 4.06 (dd, *J* = 1.4, 10.0 Hz, 1H, H-6), 4.17 (ddd, *J* = 3.3, 7.0, 15.0 Hz, 1H, H-9b), 4.98 (td, *J* = 4.7, 11.4 Hz, 1H, H-4), 5.18 (dt, *J* = 3.7, 7.5 Hz, 1H, H-8), 5.61 (dd, *J* = 1.5, 7.5 Hz, 1H, H-7), 6.90 (m, 2H, NH₂), 7.43, 7.74 (AA', BB' of AA'BB', *J* = 8.6 Hz, 4H, CH_{ar}). ¹³C NMR (CD₂Cl₂) δ 11.8 (SMe), 21.3, 21.4, 21.4 (3 OAc), 37.0 (C-3), 39.4 (C-9), 52.2 (C-5), 53.3 (OMe), 69.5 (C-7), 71.3 (C-8), 73.0 (C-4), 73.8 (C-6), 84.7 (C-2), 129.1, 129.3, 130.1, 133.5 (6C, C-Ar), 166.3, 168.2, 170.3, 170.4, 171.4 (5 CO). ESI-MS calcd for C₂₄H₃₁ClN₂O₁₀S [M + Na]⁺ 597.13; found *m/z* 597.14.

Methyl (Methyl 5-Cyclopropylamido-4,7,8-tri-*O*-acetyl-9-(4-chlorobenzamido)-3,5,9-trideoxy-2-thio-*D*-glycero- α -*D*-galacto-2-nonulopyranosid)onate (22f). Compound **21** (74.0 mg, 0.16 mmol) was dissolved in dry DCM (2.0 mL) under argon atmosphere. Cyclopropanoyl chloride (24.0 μ L, 28.0 mg, 0.27 mmol), NEt₃ (37.0 μ L, 27.0 mg, 0.27 mmol), and a catalytic amount of DMAP were added successively. The reaction mixture was stirred at rt for 3.5 h. Then it was washed with saturated aqueous NaHCO₃ (3 \times 5 mL) and H₂O (1 \times 5 mL). The organic phase was dried over Na₂SO₄, filtered, and concentrated under reduced pressure. The crude product was purified by chromatography on silica gel (EA:PE, gradient 1:1 to 2:1) to yield **22f** (43 mg, 75%). [α]_D²⁰ 0.48 (*c* 2.16, CH₂Cl₂). ¹H NMR (500 MHz, CDCl₃) δ 0.70 (m, 2H, CH₂), 0.91 (m, 2H, CH₂), 1.22 (m, 1H, CH), 2.01 (m, 4H, H-3a, SMe), 2.10, 2.12, 2.20 (3 s, 9H, 3 OAc), 2.71 (dd, *J* = 4.6, 12.7 Hz, 1H, H-3b), 2.86 (m, 1H, H-9a), 3.73 (dd, *J* = 2.0, 10.7 Hz, 1H, H-6), 3.76 (s, 3H, OMe), 4.24 (q, *J* = 10.5 Hz, 1H, H-5), 4.36 (ddd, *J* = 2.9, 8.7, 11.6 Hz, 1H, H-9b), 4.84 (td, *J* = 4.6, 11.6 Hz, 1H, H-4), 5.12 (dd, *J* = 2.0, 10.1 Hz, 1H, H-7), 5.25 (dt, *J* = 2.7, 10.1 Hz, 1H, H-8), 5.42 (d, *J* = 10.3 Hz, 1H, 5-NH), 7.18 (ddd, *J* = 4.0, 8.7 Hz, 1H, NH), 7.39, 7.75 (AA', BB' of AA'BB', *J* = 8.6 Hz, 4H, CH_{ar}). ¹³C NMR (CDCl₃) δ 7.4 (2C, CH₂), 12.1 (CH), 14.6 (SMe), 21.1 (3C, 3 OAc), 37.8, 38.2 (2C, C-3, C-9), 49.4 (C-5), 52.9 (OMe), 67.9, 68.2 (2C, C-7, C-8), 69.7 (C-4), 73.9 (C-6), 82.8 (C-2), 128.8, 129.4, 132.7, 137.7 (6C, C-Ar), 166.1, 167.7, 170.3, 171.2, 172.4, 173.7 (6 CO). ESI-MS calcd for C₂₈H₃₅ClN₂O₁₁S [M + Na]⁺ 665.17; found *m/z* 665.06.

Methyl (Methyl 5-Cyclobutylamido-4,7,8-tri-*O*-acetyl-9-(4-chlorobenzamido)-3,5,9-trideoxy-2-thio-*D*-glycero- α -*D*-galacto-2-nonulopyranosid)onate (22g). Compound **21** (60 mg, 0.1 mmol) was dissolved in dry DCM (2.4 mL). Cyclobutanecarbonyl chloride (36 μ L, 37 mg, 0.3 mmol), NEt₃ (44 μ L, 32 mg, 0.3 mmol), and a catalytic amount of DMAP were added successively. The reaction mixture was stirred at rt overnight. Then it was washed with saturated aqueous NaHCO₃ (3 \times 5 mL) and H₂O (1 \times 5 mL). The organic phase was dried over Na₂SO₄, filtered, and concentrated under reduced pressure. The crude product was purified by chromatography on silica gel (EA:PE, gradient 1:1 to 2:1) to yield **22g** (27 mg, 39%). ¹H NMR (500 MHz, CDCl₃) δ 2.00 (m, 4H, SMe, H-3a), 2.03 (s, 3H, OAc), 2.08 (m, 2H, CH₂), 2.13 (s, 3H, OAc), 2.18 (m, 4H, CH₂), 2.27 (s, 3H, OAc), 2.71 (dd, *J* = 4.6, 12.7 Hz, 1H, H-3b), 2.86 (m, 2H, CH, H-9a), 3.71 (dd, *J* = 2.1, 10.8 Hz, 1H, NH), 3.77 (s, 3H, OMe), 4.20 (t, *J* = 10.4 Hz, 1H, H-5), 4.37 (ddd, *J* = 2.9, 8.8, 15.2 Hz, 1H, H-9b), 4.82 (td, *J* = 4.6, 11.6 Hz, 1H, H-4), 5.08 (d, *J* = 10.3 Hz, 1H, H-6), 5.27 (m, 2H, H-7, H-8), 7.13 (dd, *J* = 4.0, 8.6 Hz, 1H, 5-NH), 7.38, 7.74 (AA', BB' of AA'BB', *J* = 8.6 Hz, 4H, CH_{ar}). ¹³C NMR (CDCl₃) δ 12.1 (SMe), 18.2 (CH₂), 20.9 (3C, OAc), 24.9 (2C, CH₂), 37.8, 38.2 (2C, C-3, C-9), 39.8 (CH), 49.2 (C-5), 53.0 (OMe), 68.0, 68.1, 69.6 (3C, C-4, C-7, C-8), 73.8 (C-4), 82.8 (C-2), 128.4, 132.7, 137.7 (6C, C-Ar), 166.1, 167.7, 170.3, 172.6, 175.0 (6C, CO). ESI-MS calcd for C₂₉H₃₇ClN₂O₁₁S [M + Na]⁺ 679.18; found *m/z* 679.11.

Methyl ((2,3-Difluorobenzyl) 5-Cyclopropylamido-4,7,8-tri-*O*-acetyl-9-(4-chlorobenzamido)-3,5,9-trideoxy-*D*-glycero- α -*D*-galacto-2-nonulopyranosid)onate (23f). Compound **22f** (37.0 mg, 0.06 mmol) was reacted with 2,3-difluorobenzyl alcohol (18.0 μ L, 23.0 mg, 0.16 mmol), *N*-iodosuccinimide (20.0 mg, 0.09 mmol), and triflic acid (4.00 μ L, 7.00 mg, 0.05 mmol). The crude product was purified by chromatography on silica gel (1% gradient *i*PrOH in petrol ether/DCM 2:1) to yield **23f** (31 mg, 72%). [α]_D²⁰ 0.23 (*c* 1.73, CH₂Cl₂). ¹H NMR (500 MHz, CDCl₃) δ 0.79 (m, 2H, CH₂), 1.00 (m, 2H, CH₂), 1.29 (m, 1H, CH), 2.08 (m, 1H, H-3a), 2.09, 2.21, 2.29 (3 s, 9H, 3 OAc), 2.72 (dd, *J* = 4.5, 12.8 Hz, 1H, H-3b), 2.98 (m, 1H, H-9a), 3.82 (s, 3H, OMe), 4.11 (m, 1H, H-6), 4.34 (m, 1H, H-5), 4.41 (m, 1H, H-9b), 4.59, 4.89 (A, B of AB, *J* = 12.0 Hz, 2H, CH₂Ar), 4.95 (m, 1H, H-4), 5.21 (dd, *J* = 1.8, 10.0 Hz, 1H, H-7), 5.36 (m, 1H, H-8), 5.49 (d, *J* = 10.2 Hz, 1H, 5-NH), 7.19 (m, 4H, NH, CH_{ar}), 7.47, 7.84 (AA', BB', *J* = 8.5 Hz, 4H, CH_{ar}). ¹³C NMR (CDCl₃) δ 7.1, 7.4 (2C, 2 CH₂), 14.6 (SMe), 20.8, 21.2, 21.3 (3 OAc), 25.4 (CH), 38.0, 38.4 (2C, C-3, C-9), 49.5 (C-5), 52.8 (OMe), 60.5 (CH₂Ar), 67.9, 68.4, 68.9 (3C, C-4, C-7, C-8), 72.5 (C-6), 98.6 (C-2), 115.8, 117.1, 123.9, 125.4, 126.7, 128.5, 132.7, 137.7 (12C, C-Ar), 166.2, 168.0, 170.3, 171.2, 172.4, 173.8 (6 CO). ESI-MS calcd for C₃₄H₃₇ClF₂N₂O₁₂[M + Na]⁺ 761.20; found *m/z* 761.16.

Methyl ((2,3-Difluorobenzyl) 5-Cyclobutylamido-4,7,8-tri-*O*-acetyl-9-(4-chlorobenzamido)-3,5,9-trideoxy-*D*-glycero- α -*D*-galacto-2-nonulopyranosid)onate (23g). Compound **22g** (53.0 mg, 0.08 mmol) was reacted with 2,3-difluorobenzyl alcohol (25.0 μ L, 33.0 mg, 0.23 mmol), *N*-iodosuccinimide (27.0 mg, 0.12 mmol), and triflic acid (6.00 μ L, 10.0 mg, 0.07 mmol). The crude product was purified by chromatography on silica gel (1% gradient *i*PrOH in petrol ether/DCM 2:1) to yield **23g** (44 mg, 72%). [α]_D²⁰ 0.23 (*c* 1.7, CH₂Cl₂). ¹H NMR (500 MHz, CDCl₃) δ 1.85 (m, 1H, H-3a), 1.99 (s, 3H, OAc), 2.04 (m, 4H, 2 CH₂), 2.13 (s, 3H, OAc), 2.18 (m, 2H, CH₂), 2.27 (s, 3H, OAc), 2.63 (dd, *J* = 4.5, 12.7 Hz, 1H, H-3b), 2.84 (quint, *J* = 10.0 Hz, 1H, CH), 2.91 (dt, *J* = 3.5, 15.0 Hz, 1H, H-9a), 3.73 (s, 3H, OMe), 4.03 (d, *J* = 12.5 Hz, 1H, H-6), 4.22 (q, *J* = 10.4 Hz, 1H, H-5), 4.33 (ddd, *J* = 2.8, 8.6, 15.1 Hz, 1H, H-9b), 4.51 (A of AB, *J* = 11.8 Hz, 1H, CH₂Ar), 4.82 (m, 2H, H-4, CH₂Ar), 5.10 (dd, *J* = 1.7, 9.9 Hz, 1H, H-7), 5.17 (d, *J* = 10.4 Hz, 1H, 5-NH), 5.26 (m, 1H, H-8), 5.36 (d, *J* = 10.2 Hz, 1H, 5-NH), 7.09 (m, 3H, CH_{ar}), 7.19 (dd, *J* = 4.9, 8.3 Hz, 1H, NH), 7.37, 7.73 (AA', BB' of AA'BB', *J* = 8.4 Hz, 4H, CH_{ar}). ¹³C NMR (CDCl₃) δ 18.1 (2C, CH₂), 21.2 (3C, OAc), 25.3 (CH₂), 37.9 (C-3), 38.3 (C-9), 39.7 (CH), 49.1 (C-5), 52.8 (OMe), 60.3 (CH-Ar), 67.8, 68.2, 68.8 (3C, C-4, C-7, C-8), 72.3 (C-6), 98.5 (C-2), 116.4, 117.1, 123.7, 125.3, 128.7, 132.5, 137.6 (12C, C-Ar), 167.9, 171.1, 172.4, 175.0 (6C, CO). ESI-MS calcd for C₃₅H₃₉ClF₂N₂O₁₂ [M + Na]⁺ 775.22; found *m/z* 775.25.

Hapten Inhibition Assays with MAG_{d1-3}-Fc. Murine MAG_{d1-3}-Fc was affinity purified from CHO-Lec 3.2.8.1 cell culture supernatant as described before,⁴⁰ dialyzed against 10 mM phosphate buffer pH 7.4, sterile filtered, and stored at 4 °C. The purified protein is stable for several months. The protein was analyzed by an ELISA and binding assay with immobilized fetuin. Inhibition assays for MAG were performed as described previously.^{26,30,40} In brief, fetuin was immobilized in microtiter plates and binding of MAG-Fc was determined in the presence of seven to eight different concentrations for each inhibitor using alkaline phosphatase-labeled anti-Fc antibodies. The half-maximal inhibitory concentrations were determined from corresponding binding curves and used to calculate relative inhibitory concentrations (rIC₅₀).

SPR Analysis. The SPR measurements were performed on a Biacore 3000 surface plasmon resonance based optical biosensor (Biacore AB, Sweden). Sensor chips (CM5 and CM4), immobilization kits, maintenance supply and HBS-EP (10 mM HEPES pH 7.4, 150 mM NaCl, 3 mM EDTA, 0.005% v/v surfactant P20) were purchased from Biacore AB (HBS-EP ready-to-use; degassed and filtered). CM5 (CM4, respectively) chips were preconditioned prior to usage by injecting a series of

conditioning solutions. A flow rate of 50 $\mu\text{L}/\text{min}$ was used and $2 \times 20 \mu\text{L}$ of 50 mM NaOH, 10 mM HCl, 0.1% SDS, and 100 mM H_3PO_4 were injected. The carboxy groups on the CM5 (CM4) chip were activated for 10 min with a 1:1 mixture of 0.1 M *N*-hydroxysuccinimide (NHS) and 0.1 M 3-(*N,N*-dimethylamino)propyl-*N*-ethylcarbodiimide (EDC) at a flow rate of 10 $\mu\text{L}/\text{min}$. Protein A (P6031) was purchased from Sigma. A sample and a reference surface were prepared sequentially or in parallel. For immobilizing protein A, a stock solution (1 mg/mL in 50 mM phosphate buffer, pH 7.0) was diluted in 10 mM sodium acetate, pH 5.0 to obtain a concentration of 30 $\mu\text{g}/\text{mL}$. This solution was then injected over the activated surface for 10 min at a flow rate of 10 $\mu\text{L}/\text{min}$. Protein A densities around 4000 RU and 5000 RU were achieved. Flow cells were blocked with a 10 min injection of 1 M ethanolamine, pH 8.0. For capturing, $\text{MAG}_{\text{d}1-3}\text{-Fc}$ solution (expressed and purified as described⁴⁰) was diluted to a 30–40 $\mu\text{g}/\text{mL}$ concentration using HBS-EP. Afterward, $\text{MAG}_{\text{d}1-3}\text{-Fc}$ was injected at a flow rate of 1 $\mu\text{L}/\text{min}$ for 10 min. The surface was equilibrated overnight at a flow rate of 5 $\mu\text{L}/\text{min}$, achieving densities around 2000 RU. 10-fold dilution series were freshly prepared in eluent buffer immediately before use. All binding experiments were conducted at 25 °C (except thermodynamic measurements) at a flow rate of 20 $\mu\text{L}/\text{min}$. The samples were injected over 1 min, followed by 1 min dissociation. Each sample was measured with a duplicate of one concentration, using a randomized concentration order. Several buffer samples were injected before the first concentration, and one blank between each concentration, which were used for the double blank referencing during data processing. Double referencing was applied to correct for bulk effects and other systematic artifacts. Data processing and equilibrium binding constant determinations were accomplished with Scrubber (BioLogic Software, Version 1.1 g or 2.0a). Kinetic data were simultaneously fit using the nonlinear regression program Clamp or Scrubber 2.0a.

Stability Test. To determine the stability of a compound in the central nervous system, an artificial cerebrospinal fluid (aCSF) was prepared based on published data.⁴¹ The following concentrations (all in mM/L) were used: sodium 140, chlorine 125, hydrogen carbonate 22.5, potassium 2.9, calcium 1.15, magnesium 1, urea 4.16, and glucose 3.2. Because the composition of proteins in the CSF is comparable to the serum but at a lower concentration (the ratio of liquor protein to serum protein is 4×10^{-3}),⁵⁹ 0.4% v/v of human plasma (Sigma-Aldrich) was added and the pH was adjusted to 7.3. Then 100 μM solutions of the compound were prepared and shaken at 37 °C and 300 rpm on an Eppendorf-Thermomixer Comfort. Samples were withdrawn after 0, 30, 60, 120, 180 min and 20 h, respectively. The value assigned to every time point was the average of a triplicate measurement. The quantification of the samples was performed on a Agilent 1100 series HPLC instrument with a UV-DAD spectrometer using the ChemStation software.

logD_{7.3} Determination. Two similar ratios of octanol to buffer were chosen according to the expected logD value, whereas every ratio was measured as a triplicate. Phosphate buffer at pH 7.3 was prepared and shaken overnight together with octanol in order to mutually saturate the two phases. Upon separation of the two layers, the buffer phase was withdrawn and mixed with an analyte stock solution in DMSO to yield a final concentration of 10^{-4} M. Both phases were transferred to a PCR plate, which was covered with aluminum foil (Axygen PCR-AS-200) and shaken for 1 h at 1200 rpm and 25 °C on a PHMP-4 instrument (Grant-bio). After 2.5 h at room temperature, the aqueous phases were transferred to microvials, centrifuged for 30 s, and analyzed by HPLC (Agilent 1100 series). The values were accepted if the mean values of the two ratios did not differ by more than 0.1 unit.

BBB-PAMPA.⁴² Consumables (system solution, P/N 110151; brain sink buffer, P/N 110674; BBB-1 lipid solution, P/N 110672; preloaded PAMPA sandwich with stirring devices,

P/N 110212) were purchased from pION. Each donor compartment of the preloaded PAMPA plate was filled with 200 μL of pION's system solution at pH 7.4, containing the analytes at a concentration of 50 μM . Then 150 μL of the same solution were transferred to an UV-plate (UV-Star, Greiner Bio-one) and UV spectra were recorded as reference on a SpectraMax instrument (Molecular Devices). The filter membranes of the acceptor compartments were impregnated with 5 μL of BBB-1 lipid solution and each compartment was filled with 200 μL of brain sink buffer. The system was assembled and individual stirring of the wells was induced by pION's GutBox to yield an unstirred water layer thickness of 40 μm . After 30 min, the UV data of the acceptor and the donor plate were acquired on a SpectraMax instrument (Molecular Devices) and analyzed by the PAMPA Evolution command software (version 3.4, pION).

NMR. Shigemi NMR tubes were used to reduce the sample volume needed for measurement to 250 μL . $\text{MAG}_{\text{d}1-3}\text{-Fc}$ protein was diluted from a stock solution of 1 mg/mL by a factor of 2 using 99.8% D_2O (Armar Chemicals). Following dilution, the 0.5 mg/mL $\text{MAG}_{\text{d}1-3}\text{-Fc}$ was in a solvent of 50% D_2O and 50% H_2O , with 0.01% NaN_3 with a buffer of 5 mM PBS. Stock solutions of **4** were prepared in D_2O at 100, 50, and 20 mM and added to the NMR samples containing $\text{MAG}_{\text{d}1-3}\text{-Fc}$ for both the titration curve and competition experiments. Stock solutions of **13f** and **25** were prepared in D_2O at 5 mM to add to the NMR samples containing $\text{MAG}_{\text{d}1-3}\text{-Fc}$ for the competition experiments. All NMR experiments were carried out at 300 K on a Bruker DRX500 spectrometer, equipped with Z-gradient SEI probe. The pulse sequence used for the selective inversion recovery experiments began with a selective 25 ms I-Burp-1⁶⁰ 180° pulse applied to the *para*-hydrogen of the benzamide group of compound **4**. This proton does not overlap with any other resonances of **13f** and **25**. A further benefit of the *para*-hydrogen of the benzamide group of compound **4** was that its resonance frequency was sufficiently different from the water resonance, thus avoiding complications due to radiation damping.⁶¹ Following the selective inversion pulse, a 1 ms gradient pulse was applied to dephase any residual transverse magnetization. The gradient pulse was followed by a variable delay to allow for the recovery of longitudinal magnetization. The delay was followed by a DPF-GSE water suppression sequence to suppress the magnetization from the 50% H_2O .⁶¹

For each selective inversion recovery time measurement (sT1), 10 experiments were performed. These experiments consisted of increasing delays following the selective inversion pulse and gradient of 0.1, 0.25, 0.5, 0.75, 1, 1.5, 2, 3, 5, and 10 s. Then 32 scans, preceded by eight dummy scans, were measured for the concentrations of compound **4** of 500 μM and 1 mM. Sixteen scans, preceded by eight dummy scans, were measured for the concentrations of compound **4** of 2, 4, and 7 mM. For the competitive experiments with 25 μM of either **13f** or **25** added to 500 μM of **5**, 32 scans, preceded by eight dummy scans, were measured. A delay of 20 s following the measurement of each transient was inserted to allow the magnetization to return to equilibrium. Prior to the measurement upon addition of either **13f** or **25**, a 1 h incubation time for equilibration was allowed. The NMR data were analyzed using XWINNMR version 3.5 operating on a PC running under Linux OS. The spectra were apodized with an exponential decay function with 2 Hz line broadening. The inversion recovery data, as well as the one-site binding model, were fit using Prism 4 (GraphPad Software Inc., San Diego, CA).

Acknowledgment. We thank the Volkswagen Foundation, the Swiss National Science Foundation, the German Federal Ministry for Education and Research (BMBF, project 031632A), and the Tönjes–Vagt Foundation (project XXI) for their support of this work.

Supporting Information Available: Surface plasmon resonance assay, structure of compounds implemented additionally

in the Biacore–Hapten assay correlation, HRMS data, HPLC traces, and ^1H spectra of target compounds **13a–f**, **19a–g**. This material is available free of charge via the Internet at <http://pubs.acs.org>.

References

- Schwab, M. E.; Bandtlow, C. E. Neurobiology—Inhibitory Influences. *Nature* **1994**, *371*, 658–659.
- Ramon y Cajal, S. *Degeneration and Regeneration of the Nervous System*; Oxford University Press: London, 1928.
- Schwab, M. E.; Caroni, P. Oligodendrocytes and CNS myelin are nonpermissive substrates for neurite growth and fibroblast spreading *in vitro*. *J. Neurosci.* **1988**, *8*, 2381–2393.
- Sandvig, A.; Berry, M.; Barrett, L. B.; Butt, A.; Logan, A. Myelin-, reactive glia-, and scar-derived CNS axon growth inhibitors: expression, receptor signaling, and correlation with axon regeneration. *Glia* **2004**, *46*, 225–251.
- Filbin, M. T. Myelin-associated inhibitors of axonal regeneration in the adult mammalian CNS. *Nature Rev. Neurosci.* **2003**, *4*, 703–713.
- He, Z. G.; Koprivica, V. The Nogo signaling pathway for regeneration block. *Annu. Rev. Neurosci.* **2004**, *27*, 341–368.
- Caroni, P.; Savio, T.; Schwab, M. E. Central nervous-system regeneration—oligodendrocytes and myelin as nonpermissive substrates for neurite outgrowth. *Prog. Brain Res.* **1988**, *78*, 363–370.
- Quarles, R. H. A hypothesis about the relationship of myelin-associated glycoprotein's function in myelinated axons to its capacity to inhibit neurite outgrowth. *Neurochem. Res.* **2009**, *34*, 79–86.
- Quarles, R. H. Myelin-associated glycoprotein (MAG): past, present and beyond. *J. Neurochem.* **2007**, *100*, 1431–1448.
- Crocker, P. R.; Clark, E. A.; Filbin, M.; Gordon, S.; Jones, Y.; Kehrl, J. H.; Kelm, S.; Le Douarin, N.; Powell, L.; Roder, J.; Schnaar, R. L.; Sgroi, D. C.; Stamenkovic, K.; Schauer, R.; Schachner, M.; van den Berg, T. K.; van der Merwe, P. A.; Watt, S. M.; Varki, A. Siglecs: A Family of Sialic Acid Binding Lectins. *Glycobiology* **1998**, *8*, Glycoforum 2 v–vi.
- Kelm, S.; Pelz, A.; Schauer, R.; Filbin, M. T.; Tang, S.; de Bellard, M. E.; Schnaar, R. L.; Mahoney, J. A.; Hartnell, A.; Bradfield, P. Sialoadhesin, myelin-associated glycoprotein and CD22 define a new family of sialic acid-dependent adhesion molecules of the immunoglobulin superfamily. *Curr. Biol.* **1994**, *4*, 965–972.
- Lauren, J.; Hu, F.; Chin, J.; Liao, J.; Airaksinen, M. S.; Strittmatter, S. M. Characterization of Myelin Ligand Complexes with Neuronal Nogo-66 Receptor Family Members. *J. Biol. Chem.* **2007**, *282*, 5715–5725.
- Robak, L. A.; Venkatesh, K.; Lee, H.; Raiker, S. J.; Duan, Y.; Lee-Osbourne, J.; Hofer, T.; Mage, R. G.; Rader, C.; Giger, R. J. Molecular basis of the interactions of the Nogo-66 receptor and its homolog NgR2 with myelin-associated glycoprotein: development of NgR^{DM1}-Fc, a novel antagonist of CNS myelin inhibition. *J. Neurosci.* **2009**, *29*, 5766–5783.
- Yang, L. J. S.; Zeller, C. B.; Shaper, N. L.; Kiso, M.; Hasegawa, A.; Shapiro, R. E.; Schnaar, R. L. Gangliosides are neuronal ligands for myelin-associated glycoprotein. *Proc. Nat. Acad. Sci. U.S.A.* **1996**, *93*, 814–818.
- Tang, S.; Shen, Y. J.; DeBellard, M. E.; Mukhopadhyay, G.; Salzer, J. L.; Crocker, P. R.; Filbin, M. T. Myelin-associated glycoprotein interacts with neurons via a sialic acid binding site at ARG118 and a distinct neurite inhibition site. *J. Cell Biol.* **1997**, *138*, 1355–1366.
- Vinson, M.; Strijbos, P. J. L. M.; Rowles, A.; Facci, L.; Moore, S. E.; Simmons, D. L.; Walsh, F. S. Myelin-associated glycoprotein interacts with ganglioside GT1b—A mechanism for neurite outgrowth inhibition. *J. Biol. Chem.* **2001**, *276*, 20280–20285.
- Wörter, V.; Schweigreiter, R.; Kinzel, B.; Mueller, M.; Barske, C.; Böck, G.; Frenzel, S.; Bandtlow, C. E. Inhibitory Activity of Myelin-Associated Glycoprotein on Sensory Neurons Is Largely Independent of NgR1 and NgR2 and Resides within Ig-Like Domains 4 and 5. *PLoS One* **2009**, *4*, e2518; DOI:10.1371/journal.pone.0005218.
- Yang, L. J. S.; Lorenzini, I.; Vajn, K.; Mountney, A.; Schramm, L. P.; Schnaar, R. L. Sialidase enhances spinal axon outgrowth *in vivo*. *Proc. Nat. Acad. Sci. U.S.A.* **2006**, *103*, 11057–11062.
- Collins, B. E.; Kiso, M.; Hasegawa, A.; Tropak, M. B.; Roder, J. C.; Crocker, P. R.; Schnaar, R. L. Binding specificities of the sialoadhesin family of I-type lectins—sialic acid linkage and substructure requirements for binding of myelin-associated glycoprotein, Schwann cell myelin protein, and sialoadhesin. *J. Biol. Chem.* **1997**, *272*, 16889–16895.
- Vyas, A. A.; Blixt, O.; Paulson, J. C.; Schnaar, R. L. Potent glycan inhibitors of myelin-associated glycoprotein enhance axon outgrowth *in vitro*. *J. Biol. Chem.* **2005**, *280*, 16305–16310.
- Ito, H.; Ishida, H.; Collins, B.; Fromholt, S.; Schnaar, R.; Kiso, M. Systematic synthesis and MAG-binding activity of novel sulfated GM1b analogues as mimics of Chol-1 (alpha-series) gangliosides: highly active ligands for neural Siglecs. *Carbohydr. Res.* **2003**, *338*, 1621–1639.
- Schwardt, O.; Gaethje, H.; Vedani, A.; Mesch, S.; Gao, G.; Spreafico, M.; von Orelli, J.; Kelm, S.; Ernst, B. Examination of the Biological Role of the alpha(2–6)-Linked Sialic Acid in Gangliosides Binding to the Myelin-Associated Glycoprotein (MAG). *J. Med. Chem.* **2009**, *52*, 989–1004.
- Bhunja, A.; Schwardt, O.; Gaethje, H.; Gao, G.; Kelm, S.; Benie, A. J.; Hricovini, M.; Peters, T.; Ernst, B. Consistent Bioactive Conformation of the Neu5Ac alpha(2–3)Gal Epitope Upon Lectin Binding. *ChemBioChem* **2008**, *9*, 2941–2945.
- Shin, S.; Gäthje, H.; Schwardt, O.; Gao, G.; Ernst, B.; Kelm, S.; Meyer, B. Binding epitopes of gangliosides to their neuronal receptor, myelin-associated glycoprotein, from saturation transfer difference NMR. *ChemBioChem* **2008**, *9*, 2946–2949.
- Schwizer, D.; Gäthje, H.; Kelm, S.; Porro, M.; Schwardt, O.; Ernst, B. Antagonists of the myelin-associated glycoprotein: a new class of tetrasaccharide mimics. *Bioorg. Med. Chem.* **2006**, *14*, 4944–4957.
- Kelm, S.; Brossmer, R.; Isecke, R.; Gross, H. J.; Streng, K.; Schauer, R. Functional groups of sialic acids involved in binding to Siglecs (Sialoadhesins) deduced from interactions with synthetic analogues. *Eur. J. Biochem.* **1998**, *255*, 663–672.
- Kelm, S.; Brossmer, R. Neuraminic acid derivatives for use as Siglec inhibitors. PCT Patent WO 03/000709A2, **2003**.
- Streng, K.; Schauer, R.; Bovin, N.; Hasegawa, A.; Kiso, M.; Kelm, S. Glycan specificity of myelin-associated glycoprotein and sialoadhesin deduced from interactions with synthetic oligosaccharides. *Eur. J. Biochem.* **1998**, *258*, 677–685.
- Gao, G.; Smiesko, M.; Schwardt, O.; Gäthje, H.; Kelm, S.; Vedani, A.; Ernst, B. Mimetics of the tri- and tetrasaccharide epitope of GQ1b/alpha as myelin-associated glycoprotein (MAG) ligands. *Bioorg. Med. Chem.* **2007**, *15*, 7459–7469.
- Shelke, S. V.; Gao, G. P.; Mesch, S.; Gäthje, H.; Kelm, S.; Schwardt, O.; Ernst, B. Synthesis of sialic acid derivatives as ligands for the myelin-associated glycoprotein (MAG). *Bioorg. Med. Chem.* **2007**, *15*, 4951–4965.
- Hasegawa, A.; Ohki, H.; Nagahama, T.; Ishida, H.; Kiso, M. A facile, large-scale preparation of the methyl 2-thioglycoside of N-acetylneuraminic acid, and its usefulness for the alpha-steroselective synthesis of sialoglycosides. *Carbohydr. Res.* **1991**, *212*, 277–281.
- Fitz, W.; Rosenthal, P. B.; Wong, C. H. Synthesis and inhibitory properties of a thiomethylmercuric sialic acid with application to the X-ray structure determination of 9-O-acetylsialic acid esterase from influenza C virus. *Bioorg. Med. Chem.* **1996**, *4*, 1349–1353.
- Hossain, N.; Zapata, A.; Wilstermann, M.; Nilsson, U. J.; Magnusson, G. Synthesis of GD3-lactam: a potential ligand for the development of an anti-melanoma vaccine. *Carbohydr. Res.* **2002**, *337*, 569–580.
- Ren, C. T.; Chen, C. S.; Wu, S. H. Synthesis of a sialic acid dimer derivative, 2'-alpha-O-benzyl Neu5Ac-alpha-(2–5)Neu5Gc. *J. Org. Chem.* **2002**, *67*, 1376–1379.
- Boullanger, P.; Maunier, V.; Lafont, D. Syntheses of amphiphilic glycosylamides from glycosyl azides without transient reduction to glycosylamines. *Carbohydr. Res.* **2000**, *324*, 97–106.
- Alzuet, G.; Ferrer, S.; Borrás, J.; Solans, X. Structure of methazolamide—an inhibitor of carbonic anhydrase. *Acta Crystallogr., Sect. C: Cryst. Struct. Commun.* **1991**, *47*, 2377–2379.
- Izumi, M.; Shen, G. J.; Wacowich-Sgarbi, S.; Nakatani, T.; Plettenburg, O.; Wong, C. H. Microbial glycosyltransferases for carbohydrate synthesis: alpha-2,3-sialyltransferase from *Neisseria gonorrhoeae*. *J. Am. Chem. Soc.* **2001**, *123*, 10909–10918.
- Grehn, L.; Gunnarsson, K.; Ragnarsson, U. Removal of formyl, acetyl, and benzoyl groups from amides with conversion into the corresponding *tert*-butyl carbamates. *J. Chem. Soc., Chem. Commun.* **1985**, 1317–1318.
- Kaiser, E.; Tam, J. P.; Kubiak, T. M.; Merrifield, R. B. Chlorotrimethylsilane-phenol as a mild deprotection reagent for the *tert*-butyl based protecting groups in peptide-synthesis. *Tetrahedron Lett.* **1988**, *29*, 303–306.
- Bock, N.; Kelm, S. Binding and Inhibition Assay for Siglecs. In *Methods Mol. Biol.*, Vol. 347; Brockhausen, I., Eds.; Humana Press Inc: Totowa, NJ, 2006; pp 359–376.
- Mitchell, R.; Herbert, D.; Carman, C. Acid-base constants and temperature coefficients for cerebrospinal fluid. *J. Appl. Physiol.* **1965**, *20*, 27–30.

- (42) Di, L.; Kerns, E.; Fan, K.; McConnell, O.; Carter, G. High throughput artificial membrane permeability assay for blood-brain barrier. *Eur. J. Med. Chem.* **2003**, *38*, 223–232.
- (43) Rich, R.; Myszka, D. Advances in surface plasmon resonance biosensor analysis. *Curr. Opin. Biotechnol.* **2000**, *11*, 54–61.
- (44) Morton, T.; Myszka, D. Kinetic analysis of macromolecular interactions using surface plasmon resonance biosensors. *Methods Enzymol.* **1998**, *295*, 268–294.
- (45) Nagata, K.; Handa, H. *Real-Time Analysis of Biomolecular Interactions: Applications of Biacore*; Springer Verlag: Berlin, 2000.
- (46) Mannen, T.; Yamaguchi, S.; Honda, J.; Sugimoto, S.; Kitayama, A.; Nagamune, T. Observation of charge state and conformational change in immobilized protein using surface plasmon resonance sensor. *Anal. Biochem.* **2001**, *293*, 185–193.
- (47) Gestwicki, J. E.; Hsieh, H. V.; Pitner, J. B. Using receptor conformational change to detect low molecular weight analytes by surface plasmon resonance. *Anal. Chem.* **2001**, *73*, 5732–5737.
- (48) Stokmaier, D.; Khorev, O.; Cutting, B.; Born, R.; Ricklin, D.; Ernst, T. O. G.; Böni, F.; Schwingruber, K.; Gentner, M.; Wittwer, M.; Spreafico, M.; Vedani, A.; Rabbani, S.; Schwardt, O.; Ernst, B. Design, synthesis and evaluation of monovalent ligands for the asialoglycoprotein receptor (ASGP-R). *Bioorg. Med. Chem.* **2009**, *17*, 7254–7264.
- (49) Dalvit, C.; Flocco, M.; Knapp, S.; Mostardini, M.; Perego, R.; Stockman, B. J.; Veronesi, M.; Varasi, M. High-throughput NMR-based screening with competition binding experiments. *J. Am. Chem. Soc.* **2002**, *124*, 7702–7709.
- (50) Roos, H.; Karlsson, R.; Nilshans, H.; Persson, A. Thermodynamic analysis of protein interactions with biosensor technology. *J. Mol. Recognit.* **1998**, *11*, 204–210.
- (51) Humphrey, W.; Dalke, A.; Schulten, K. VMD—Visual Molecular Dynamics. *J. Mol. Graphics* **1996**, *14*, 33–38.
- (52) May, A. P.; Robinson, R. C.; Vinson, M.; Crocker, P. R.; Jones, E. Y. Crystal structure of the N-terminal domain of sialoadhesin in complex with 3' sialyllactose at 1.85 angstrom resolution. *Mol. Cell* **1998**, *1*, 719–728.
- (53) Crocker, R. P.; Klem, S. *The Siglec Family of I-type Lectins*; Wiley-VCH: Weinheim, 2000; Vol. IV, pp 579–595.
- (54) Sinnokrot, M. O.; Sherrill, C. D. Substituent effects in pi-pi interactions: sandwich and T-shaped configurations. *J. Am. Chem. Soc.* **2004**, *126*, 7690–7697.
- (55) Mehta, P.; Cummings, R. D.; McEver, R. P. Affinity and kinetic analysis of P-selectin binding to P-selectin glycoprotein ligand-1. *J. Biol. Chem.* **1998**, *273*, 32506–32513.
- (56) Wild, M. K.; Huang, M.-C.; Schulze-Horsel, U.; van der Merwe, P. A.; Vestweber, D. Affinity, kinetics and thermodynamics of E-selectin binding to E-selectin ligand-1. *J. Biol. Chem.* **2001**, *276*, 31602–31612.
- (57) Nicholson, M. W.; Barclay, A. N.; Singer, M. S.; Rosen, S. D.; van der Merwe, P. A. Affinity and kinetic analysis of L-selectin (CD62L) binding to glycosylation-dependent cell-adhesion molecule-1. *J. Biol. Chem.* **1998**, *273*, 763–770.
- (58) Herfurth, L.; Ernst, B.; Wagner, B.; Ricklin, D.; Strasser, D. S.; Magnani, J. L.; Benie, A. J.; Peters, T. Comparative epitope mapping with saturation transfer difference NMR of sialyl Lewis^x compounds and derivatives bound to a monoclonal antibody. *J. Med. Chem.* **2005**, *48*, 6879–6886.
- (59) Wissenschaftliche Tabellen Geigy, 8th ed.; Ciba Geigy: Basel, 1977; Vol. Teilband Körperflüssigkeiten, pp 161–173.
- (60) Geen, H.; Freeman, R. Band-selective radiofrequency pulses. *J. Magn. Reson.* **1991**, *93*, 93–141.
- (61) Cutting, B.; Chen, J. H.; Moskau, D.; Bodenhausen, G. Radiation damping compensation of selective pulses in water-protein exchange spectroscopy. *J. Biomol. NMR* **2000**, *17*, 323–330.

Appendix 5: J. Chem. Inf. Mod. 2011

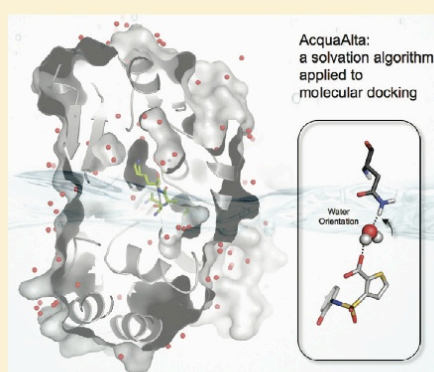
AcquaAlta: A Directional Approach to the Solvation of Ligand–Protein Complexes

Gianluca Rossato,[†] Beat Ernst,[†] Angelo Vedani,[†] and Martin Smiesko^{*,†}

[†]Institute of Molecular Pharmacy, Pharmacenter, University of Basel, Klingelbergstrasse 50, 4056 Basel, Switzerland

 Supporting Information

ABSTRACT: Water molecules mediating polar interactions in ligand–protein complexes can substantially contribute to binding affinity and specificity. To account for such water molecules in computer-aided drug design, we performed an extensive search in the Cambridge Structural Database (CSD) to identify the geometrical criteria defining interactions of water molecules with ligand and protein. In addition, with *ab initio* calculations the propensity of ligand hydration was evaluated. Based on this information, we developed an algorithm (AcquaAlta) to reproduce water molecules bridging polar interactions between ligand and protein moieties. This approach was validated with 20 crystal structures and yielded a match of 76% between experimental and calculated water positions. When water molecules establishing only weak interactions with the protein were neglected, the match could be improved to 88%. Supported by a pharmacophore-based alignment tool, the solvation algorithm was then applied to the docking of oligopeptides to the periplasmic oligopeptide binding protein A (OppA). Calculated waters based on the crystal poses matched an average of 66% of the experimental waters. With water molecules calculated based on the docked ligands, the average match with the experimental waters dropped to 53%.



■ INTRODUCTION

Water molecules mediating ligand–protein interactions can affect ligand affinity as well as increase the specificity of the interaction (i) by forming bridging hydrogen bonds¹ or (ii) by increasing the flexibility of the binding site.² Water molecules can form both intramolecular (i.e. waters between two ligand atoms or two protein atoms) and intermolecular bridges (i.e., waters between ligand and protein atoms). Hereafter, bridging water molecules will refer to solvent molecules linking polar functional groups of ligand and protein.

An analysis of 392 high-resolution complexes³ retrieved from the Protein Data Bank (PDB)⁴ showed that over 85% of the ligand–protein complexes include at least one water molecule bridging ligand and protein. In the analyzed complexes, the average number of ligand-bound water molecules was found to be 4.6. In addition, 76% of these water molecules were identified to participate in polar interactions with both ligand and protein. A statistical analysis on protein–protein complexes revealed that water-mediated interactions are as abundant as direct protein–protein hydrogen bonds.⁵

Previous studies reported the utmost importance of considering water molecules in structure-based drug design and pharmacophore modeling.^{6–8} The displacement of a water molecule can substantially affect the binding free energy. In the case of the HIV protease,⁹ the displacement of ordered water molecules by cyclic

urea inhibitors led to a substantial gain in entropy.¹⁰ On the contrary, in the case of the periplasmic oligopeptide binding protein A (OppA), an analogous displacement of water molecules decreased the binding affinity.¹¹ Consequently, in computer-aided drug discovery, substantial efforts are being taken to distinguish relevant and retainable water molecules from displaceable ones. Such a classification can be based on structural features such as hydrogen-bond counting,¹² the solvent-accessible polar surface area, the polarity of the withholding cavity, the characterization of the binding-pocket shape,¹³ and water conservation in homologous proteins,¹⁴ as well as on advanced statistical methods.¹⁵ Water molecules located in polar cavities forming two or more hydrogen bonds are typically considered structural waters.^{3,16} Such waters are therefore frequently conserved in protein complexes with different ligands and show thermal factors (B_{iso}) in the same range as the protein atoms they are hydrogen-bonded to.³ As a thermodynamic consequence, binding energies of ligands displacing structural water molecules suffer from a loss in enthalpy, which may only be partially compensated by a gain in entropy.

Despite previous attempts using a hydration penalty score¹⁷ or defining general rules for the treatment of water molecules

Received: March 29, 2011

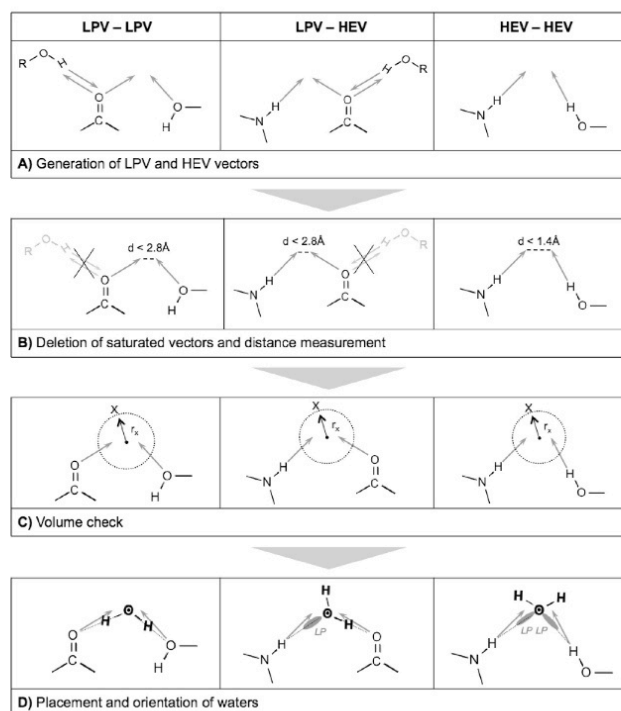


Figure 1. Figure displaying different steps of the vectors generation, scans, placement, and orientation procedure for the possible kinds of vectors combination (LPV–LPV, LPV–HEV, HEV–HEV): A) generation of the LPV and HEV vectors, B) scan for vectors not involved in other inter- or intramolecular hydrogen bonds and measurement of the distance to the unsaturated vectors; C) check of availability of space through distance measurement toward ligand and protein atoms (X) where r_v is the van der Waals radius of the atom +0.6 Å; D) placement and orientation of a water molecule toward the atoms origin of the vectors involved in the hydrogen bond.

binding at ligand–protein interfaces,¹⁸ a thorough and general parametrization is difficult to extrapolate. Therefore, the solvation pattern of the binding site at each target protein should be analyzed individually¹⁹ with particular attention to enthalpic and entropic contributions toward binding.²⁰

Computational algorithms in molecular simulations use different approaches to hydration. Some concepts calculate free-energy changes characterizing the water contribution toward the binding affinity^{21,22} or their relation to the hydration of binding sites.^{15,23–25} Potential hydration sites can be evaluated by solvent mapping or by modular neural networks.^{26,27} Furthermore, predictive tools to characterize and distinguish roles of water in biomolecules are available.^{28–30} Hydration of molecular complexes is generally addressed either considering water implicitly³¹ or explicitly.^{32–35} Some studies emphasize the need of water molecules to improve accuracy in molecular docking,^{36–39} while others claim that the presence of water improved the docking results only marginally.⁴⁰ However, since water molecules at binding interfaces are frequently observed, their presence along with their role must be considered in any molecular simulation.

In this account, we present a novel approach (AcquaAlta) where explicit water molecules are generated at the ligand–protein interface. The underlying algorithm relies on preferred positions and orientations of water molecules as extracted from

structural information collected from the Cambridge Structural Database — CSD⁴¹ (currently including the structures of more than 500,000 organic molecules). Specifically, we searched for water molecules interacting with generic functional groups (e.g., the carbonyl query is comprehensive of all aldehydes, ketones, carboxylic acids, esters, and amides) of small organic molecules. To establish a hydration-propensity ranking, water interaction energies were obtained from *ab initio* calculations on hydrated functional groups. In AcquaAlta, water molecules bridging interactions between ligand and protein partners are generated considering the hydration propensities of the involved functional groups and aromatic moieties.

For the validation of AcquaAlta, we attempted to reproduce bridging water molecules found in 20 mostly high-resolution crystal structures. This solvation approach was then applied to the docking of oligopeptides binding to OppA. We selected OppA because a large number of high-resolution ligand–protein complexes is available. Interestingly, different hydration patterns among apo OppA and ligand-bound holo OppA have been reported.^{42,43}

The envisioned application area for AcquaAlta is ligand docking where generated water molecules (in the binding site) should increase the probability of finding the bioactive pose and therefore facilitate a more reliable ranking and scoring of the docked poses.

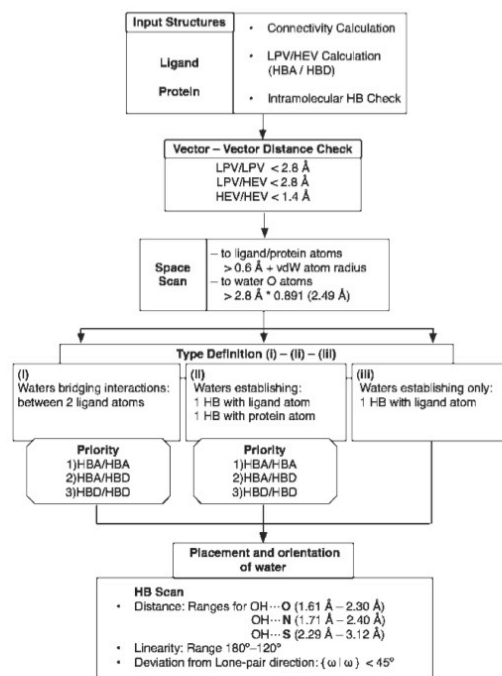


Figure 2. Flowchart representing the algorithm organization. Calculations and procedures are performed in the same order as shown in the flowchart from the top to the bottom and from left to right. HBA stands for hydrogen bond acceptors and HBD stands for hydrogen bond donors. Criteria and ranges for hydrogen bond (HB) acceptance are included.

METHODS

Cambridge Structural Database (CSD) Search. The analysis of small-molecule crystal structures regarding hydrogen bond geometries (distances, angles and torsions) with water molecules was performed in the CSD (v5.31 – November 2009 + 3 updates).⁴¹ The queries (cf. Results section) were submitted using ConQuest, and the histograms in the Supporting Information were plotted using Vista, included in the CSD software suite. The following filters were applied in the search of crystal structures of small molecules:

- All 3D coordinates determined (no calculated atom positions)
- Crystallographic R factor ≤ 0.05
- No disordered, or polymeric structures, no metal ions present
- Solely organic compounds (no organometallic compounds)

Terminal hydrogen positions were normalized during the search. Hits refer to structures matching the query search. Multiple fragments (e.g., two carboxylates interacting with water molecule(s) in the same crystal structure) for one hit were possible. In order to produce the 3D plots (see, for example, Table 3 and Table 4) we wrote a program in C++ for transforming bond distances, bond angles, and torsion angles into Cartesian coordinates.

Ab Initio Analysis of Ligand–Water Interactions. Preferred interaction geometries and associated energies of generic

functional groups of the ligand with a single water molecule were obtained from *ab initio* calculations using Gaussian 03.⁴⁴ The molecular complexes as well as the isolated entities were fully optimized in the gas phase using the Møller–Plesset perturbation theory (with energy corrections truncated at the second order, MP2) with 6-311++G(d,p) basis set.^{45,46} At the optimized geometries, an analysis of the vibrational modes was performed to confirm that they represent true minima (without imaginary frequencies detected). The final interaction energy, ΔE , was calculated as the difference between the energy of the complex and the sum of energies of its respective isolated entities. Finally, ΔE was corrected for the superposition error of the basis set using the counterpoise method.^{47,48} The generic functional groups were then ranked according to their hydration propensity based on the interaction energy (cf. Figure 4).

The setup with only a single water molecule in the gas phase was chosen to obtain an unbiased comparison of the strength of interaction (ΔE) with a given functional group for the hydration propensity scale. In reality the hydrogen-bond strength and the position of the water are also affected by nearby solvent, ligand or protein atoms.

AcquaAlta Organization. The solvation of ligand–protein complexes can be done using two different concepts: a) the solvation is performed before the docking and b) the ligand is docked and then the complex is solvated. In the former case waters are generated and oriented based solely on either ligand or protein atom positions. In the latter case (in AcquaAlta) the waters are generated and oriented based on the positions of both ligand and protein atoms simultaneously, which should lead to a better identification of waters bridging ligand–protein interactions.

Using a concept similar to the solvation module of Yeti,⁴⁹ vectors originating from hydrogen-bond donors (marking the ideal position of a hydrogen bond acceptor and termed hydrogen extension vectors, HEVs) and hydrogen-bond acceptors (marking the ideal position of a hydrogen-bond donor; lone pair vectors, LPVs) are calculated to identify ideal positions for water molecules. These vectors are generated using specific geometries for each functional group based on data retrieved from the CSD (cf. Tables 3 and 4).

In 3D space, halogen atoms present both an electropositive region (also called “corona”) along the extension of the C–X (X = Cl, Br, I) bond and an electronegative region along the axis perpendicular to this bond (also called “belt”), displaying an amphoteric character with respect to polar interactions, in ligand–protein complexes.⁵⁰ It has been observed that intermolecular distances increase with polarizability of the halogen atom involved in the interaction, i.e. larger for iodine than for fluorine.⁵⁰ In AcquaAlta, halogen atoms are considered as potential hydrogen-bond acceptors and apolar hydrogen atoms as potential hydrogen-bond donors.⁵¹ For both cases, in order to identify the starting position of the search for an interaction counterpart, the vectors are considered as the extensions of the carbon–halogen and carbon–hydrogen bond, respectively.

Figure 1 depicts the concept of AcquaAlta. First, HEV and LPV vectors are generated (Figure 1A), then, those vectors associated with atoms already engaged in inter- and intramolecular hydrogen bonds are marked as “saturated” and are not further considered (Figure 1B). Distances between unsaturated vectors are calculated and considered as potential water sites if the following criteria are met: distances of LPV–LPV and LPV–HEV below 2.8 Å, which refer to an O–H...O

C

dx.doi.org/10.1021/ci200150p | J. Chem. Inf. Model. XXXX, XXX, 000–000

Table 1. List of the 20 X-ray Structures Used for Validation of AcquaAlta along with PDB Codes, Resolutions, R-Factors, and Number of Binding Site Waters

#	biological target	PDB code	res. [Å]	R-factor	# of waters
1	trypsin	2ayw	0.97	0.138	8
2	dihydrofolate reductase	3dfr	1.70	0.152	23
3	thymidin kinase	1e2k	1.70	0.209	5
4	VEGFR2	lywn	1.71	0.206	2
5	glycogen phosphorylase	1a8i	1.78	0.182	13
6	human phosphodiesterase	1xp0	1.79	0.194	1
7	beta trypsin	1bju	1.80	0.171	2
8	holo-glyceraldehyde 3P dehydrogenase	1gd1	1.80	0.177	15
9	Hsp90	1uy6	1.90	0.184	4
10	AmpC beta-lactamase	1xgj	1.97	0.168	2
11	2CDK2	2b53	2.00	0.223	3
12	ACE	1o86	2.00	0.180	6
13	COMT	1h1d	2.00	0.174	4
14	HIV-1 protease	1hpx	2.00	0.170	4
15	non-nucleoside adenosine deaminase	1ndw	2.00	0.206	6
16	ACK1	1u4d	2.10	0.205	5
17	coagulation factor XA	1f0r	2.10	0.216	2
18	thymidin kinase	1kim	2.14	0.209	3
19	EGFR	1xkk	2.40	0.209	2
20	EGFR	1m17	2.60	0.251	1

hydrogen-bond length, and HEV–HEV distance below 1.4 Å (Figure 1B).

To ensure that an appropriate free volume for a potential water molecule is available, distances from the midpoint of the two vectors tips are checked against any ligand and protein atom. The distance for acceptance must be larger than the sum of the individual van der Waals radius of the ligand or protein atom (X) plus 0.6 Å (at this cutoff value the highest number of experimental waters was matched by the algorithm).

If the space criterion is met, the water molecule is placed at the midpoint of the vector tips (Figure 1C) into an averaged position determined by the two vectors and then oriented toward its partners (Figure 1D). In the case that a water molecule is establishing just one interaction with a ligand atom, the unengaged hydrogen or oxygen atom is oriented toward the closest free protein HEV or LPV.

The detailed organization of the algorithm is shown in Figure 2.

Input Structures. The structures of ligand and protein atoms are supplied in PDB format in order to determine connectivities for all functional groups of ligand and protein and to generate vectors (HEV, LPV, and originating from halogen and apolar hydrogen atoms). The presence of vectors already engaged in inter- and intramolecular hydrogen bonds (both in ligand and protein) is checked, and associated vectors are deleted.

Vector–Vector Distance Check. Distances are checked for LPV/LPV, LPV/HEV, and HEV/HEV combinations. Distances must remain under the defined thresholds.

Space Scan. Distances toward any ligand and protein atom are calculated to ensure that a sufficiently large volume to accommodate a water molecule is available. To check whether a water molecule has been already placed in the vicinity, distances between water oxygen atoms must be larger than 2.49 Å [sum of van der Waals radii (2.8 Å) multiplied by a factor of 0.891

Table 2. List of the OppA Crystal Structures Used in the Automated Docking Procedure along with PDB Codes, Resolutions, R-Factors, Oligopeptide Sequences, and Number of Binding Site Waters

#	PDB code	resolution [Å]	R-factor	oligopeptide sequence	# of waters
1	1JET	1.2	0.229	KAK	7
2	1JEU	1.25	0.224	KEK	9
3	1JEV	1.30	0.203	KWK	6
4	1B4Z	1.75	0.179	KDK	10
5	1B5I	1.9	0.182	KNK	7
6	1B32	1.75	0.182	KMK	7
7	1B3F	1.8	0.177	KHK	7
8	1B46	1.8	0.177	KPK	6
9	1B5I	1.8	0.179	KSK	9
10	1B58	1.8	0.179	KYK	7
11	1B5J	1.8	0.182	KQK	10
12	1B9J	1.8	0.179	KLK	6
13	1QKA	1.8	0.179	KRK	6
14	1QKB	1.8	0.181	KVK	6

(threshold value below which the Leonard-Jones 6–12 potential becomes repulsive)].

Type Definition. While generating the positions of bridging water molecules, three different types, with decreasing priority, are defined: i) waters bridging interactions within two ligand atoms, ii) waters interacting with both ligand and protein, and iii) waters establishing only a single hydrogen bond with the ligand.

Water Placement and Orientation. Finally, the water molecule is placed and oriented toward the origin of the closest unsaturated vector. Vectors are considered based on their hydration propensity as ranked from the *ab initio* calculations

Table 3. Summary of the Crystallographic Data Search for Functional Groups Acting As Hydrogen Bond Acceptors (HBA)[†]

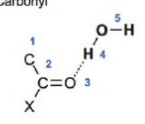
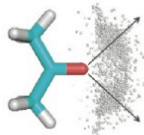
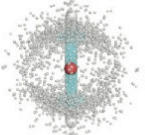
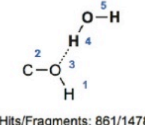


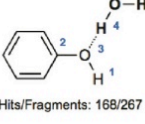
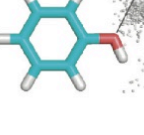

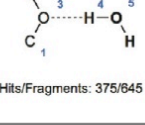
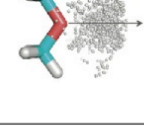
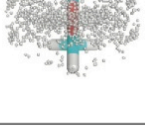
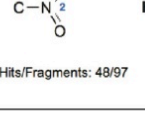
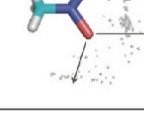
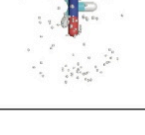
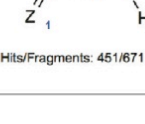


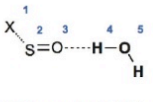
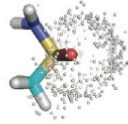
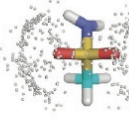
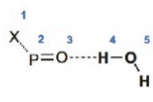


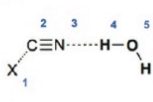


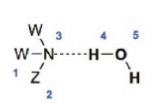
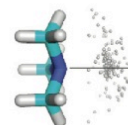
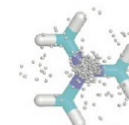
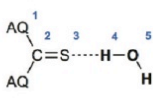
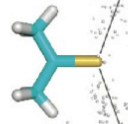
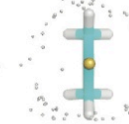
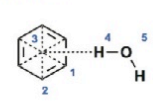
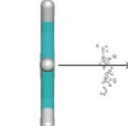
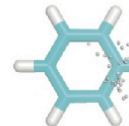
<i>HBA: Functional groups acting as hydrogen bond acceptors</i>				
<i>Entry</i>	<i>Query:</i>	<i>Side view</i>	<i>Front view</i>	<i>Measurements</i>
1	Carbonyl  Hits/Fragments: 997/1378			d [3-4] (Å) angle [3-4-5] (°) Mean 1.94±0.21 180.0±17.5 Median 1.87 165.8 LQ/HQ 1.76/2.24 134.1/175.1
2	Aliphatic hydroxyl  Hits/Fragments: 861/1478			d [3-4] (Å) angle [3-4-5] (°) Mean 2.03±0.28 151.9±23.0 Median 1.91 159.7 LQ/HQ 1.78/2.54 115.2/173.6
3	Phenolic hydroxyl  Hits/Fragments: 168/267			d [3-4] (Å) angle [3-4-5] (°) Mean 2.08±0.27 148.8±24.2 Median 1.96 156.0 LQ/HQ 1.81/2.55 112.0/172.7
4	Ether  Hits/Fragments: 375/645			d [3-4] (Å) angle [3-4-5] (°) Mean 2.24±0.31 141.4±25.1 Median 2.20 146.4 LQ/HQ 1.86/2.66 105.7/171.9
5	Nitro  Hits/Fragments: 48/97			d [3-4] (Å) angle [3-4-5] (°) Mean 2.33±0.26 112.9±23.3 Median 2.33 113.7 LQ/HQ 1.92/2.68 78.3/146.1
6	Imine  Hits/Fragments: 451/671 (***) nitrogen not charged			d [3-4] (Å) angle [3-4-5] (°) Mean 2.1±0.30 157.9±17.6 Median 1.97 164.0 LQ/HQ 1.84/2.64 130.6/175.1 (***) nitrogen not charged

Table 3. Continued

Entry	HBA: Functional groups acting as hydrogen bond acceptors			
	Query:	Side view	Front view	Measurements
7	Sulfoxide /Sulfone  Hits/Fragments: 139/271			d [3-4] (Å) angle [3-4-5] (°) Mean 1.96±0.23 159.0±16.3 Median 1.78 164.6 LQ/HQ 1.78/2.34 134.0/174.1
8	Phosphine oxide / Posphone  Hits/Fragments: 133/271			d [3-4] (Å) angle [3-4-5] (°) Mean 1.88±0.18 162.3±15.9 Median 1.84 166.4 LQ/HQ 1.73/2.0 145.5/174.9
9	Nitrile  Hits/Fragments: 30/36			d [3-4] (Å) angle [3-4-5] (°) Mean 2.10±0.23 160.4±17.4 Median 2.02 167.0 LQ/HQ 1.85/2.60 126.3/176.0
10	Tertiary amine  Hits/Fragments: 206/259			d [3-4] (Å) angle [3-4-5] (°) Mean 2.13±0.33 153.9±24.2 Median 1.97 164.3 LQ/HQ 1.83/2.7 112.8/174.4
11	Thiokethon  Hits/Fragments: 59/76			d [3-4] (Å) angle [3-4-5] (°) Mean 2.45±0.18 101.9±13.9 Median 2.40 98.9 LQ/HQ 2.26/2.70 86.9/123.0
12	Aromatic ring  Hits/Fragments: 40/43			d [3-4] (Å) angle [3-4-5] (°) Mean 2.59±0.15 86.3±12.6 Median 2.57 86.6 LQ/HQ 2.40/2.81 71.2/106.3

^a An entry number indicates each query search. In the left column, the query as submitted to the CSD; the number of hits and the number of fragments found. In the central column a typical molecule with 3D representation of the positions of water hydrogen atoms (gray spheres) and the positions of the LPVs (gray arrows). Light-blue sticks represent general fragments resembling the submitted query. In the right column, statistical results (mean, median, lower, and higher quantile) for the values of distance and linearity of hydrogen bonds between functional-group fragments and water molecules. In the 3D-plots of sulfoxide/sulfone and phosphine/phosphone the queries represent all acceptor atoms to resemble real functional groups. For clarity, in the 3D-plots of the aromatic ring as HBA, the water-hydrogen positions are displayed only for one carbon atom. In the query representation X stands for any kind of atom, Z for any kind of atom with exception of hydrogen, W for C and H atoms, AQ for C, N, O, S atoms.

(Figure 4 – Results). The geometric criteria for acceptance are loosened from very stringent to moderate giving priority to distance, linearity, and deviation from the lone-pair direction.

Predicting Bridging Water Molecules in High-Resolution Crystal Structures. Table 1 lists the proteins retrieved from the PDB for the purpose of algorithm validation. The crystal structures were selected according to their resolution, the number of bridging waters or because they were the object of previously published studies dealing with water molecules bridging polar interactions.

The PDB files were split into protein and ligand. Next, atom types based on the AMBER* force-field and atomic partial charges were assigned. Protonation and tautomeric states for ligand and protein structures were determined using Epik⁵² from the Schrodinger software suite.

The criteria defining a bridging water molecule were arbitrarily set to a distance up to 3.3 Å from any ligand atom and interaction with any kind of polar protein atom within the threshold 2.7–3.3 Å. An energy filter — a water molecule was considered “bridging”, if its total interaction energy, E_{tot} ($E_{\text{cle}} + E_{\text{vdw}} + E_{\text{HB}}$), was equal or lower than -1.0 kcal/mol after rotational optimization (oxygen atom kept fixed) based on the Yeti force field⁵³ — was applied in addition. For the comparison of experimental and calculated waters, the distance threshold for oxygen-water matching was defined as 1.4 Å (50% of the O–O distance of an OH \cdots O hydrogen bond).

Application to Docking. AcquaAlta was applied to ligand–protein complexes in which the ligand poses were generated by Alignator.⁵⁴ This pharmacophore-based alignment tool takes advantage of a conformer pool generated by a conformational search. In our case, we used MacroModel⁵² and performed in implicit water, using the OPLS2005 force-field, 20,000 search iterations (1000 steps for each rotatable bond), using an energy window of 100 kcal/mol. Each accepted conformer was then aligned to a template molecule (i.e., Lys-Asn-Lys for OppA) based on the matching pharmacophores. Of all possible solutions, only those having no close contacts with the protein and maximum number of superimposed pharmacophores were retained. Further information on this docking approach together with its application are published elsewhere.⁵⁴ Protein–ligand complexes were stripped of the solvent. Then, both crystal poses and poses produced by Alignator (14 OppA complexes – Table 2) were resolved by AcquaAlta. Finally, the water molecules were optimized using Yeti.⁵³ The deviation between crystal waters and calculated waters was monitored for each oligopeptide–protein complex using the same distance criteria as in the validation step (cf. above).

RESULTS AND DISCUSSIONS

Water Directionality. In our CSD searches we obtained detailed information on distance, linearity, and deviations from lone-pair vectors for 16 generic functional groups interacting with water molecules (Figure S1 – Supporting Information). From these data, we extracted preferred geometries for water molecules interacting with such entities for our solvation algorithm. Tables 3 and 4 list the results for the different functional groups included in the search. For each of the 16 generic functional groups the following information are shown: the query, the number of hits and fragments obtained, a 3D representation (side and front view) of water hydrogen atoms (in the case of functional groups acting as HBA) and water oxygen atoms (in the case of functional groups acting as HBD), together with relevant statistical data for distance and angle of interaction.

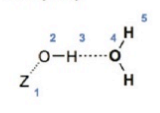

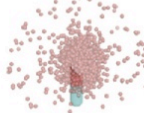
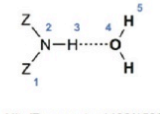
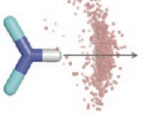
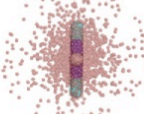
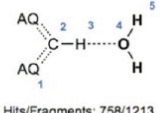
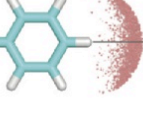
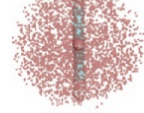
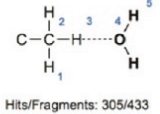


We used general queries to compile rules applicable to a wide variety of similar functional groups (e.g., carbonyl oxygen in aldehyde, ketone, carboxylic acid, ester, and amide; carboxylate groups were treated separately: cf. Figure S1 – Supporting Information) and to monitor geometries of interaction of waters involved in polar interactions. Quantitative information on specific moieties, (e.g., pyridines vs pyrimidine) and the effects of neighboring atoms of the functional groups are not extractable from the collected data (e.g., ketone vs ester).

The median length of a hydrogen bond is required to be shorter than the sum of the individual van der Waals radii. The difference between strong and weak hydrogen bonds is documented with the following examples. The mean distances for carbonyl (1.94 Å) and for phenolic hydroxyl (2.08 Å) are substantially, although not significantly shorter than the mean distances for the ones established with ether (2.24 Å) and nitro groups (2.33 Å). In addition, the spatial distribution of the water molecules interacting with a carbonyl or a phenolic/aliphatic hydroxyl positions is considerably narrowed compared to ether or nitro groups. In AcquaAlta, the acceptable lower limit for the hydrogen-bond distances is shifted compared to the values obtained from the CSD (Figure 2). This allows for a reasonable treatment of interactions between water and charged atoms where the distance is even shorter than the average hydrogen bond distance (the mean distance for the $-\text{COO}^- \cdots \text{HOH}$ interaction is 1.92 ± 0.26 Å and the mean distance for the $-\text{NR}_2\text{H}^+ \cdots \text{OH}_2$ interaction is 1.86 ± 0.21 Å).

Water prefers hydrogen bonds with a linear X-H \cdots Y arrangement. Significant deviations from this geometry are typically caused by additional interactions with nearby atoms.⁵⁵ For oxygen acceptor atoms in carbonyls, hydroxyls, sulfoxides/sulfones, phosphines/phosphones, and nitrogen acceptor atoms in tertiary amines, imines, and nitriles as well as all the hydroxyl and amino donor atoms the median values for angles vary from 156° to 180° (angle statistics in Table 3). In contrast, for acceptors such as ethers, nitro groups, thioketones, aromatic rings, and aromatic and aliphatic “hydrogen-bond donor” groups (Table 4 – entries 15, 16) the median values for angles substantially deviate from linearity, ranging from 87° (aromatic ring) to 148° (aliphatic CH). AcquaAlta allows for the X-H \cdots Y angle to range from 180° (linear) to 120°, which therefore covers most cases listed above.

Hydrogen bonds are usually directed along the lone pair(s) of the acceptor atom.⁵⁶ Since CSD/Conquest does not allow to define lone pairs, the torsion angle defined by the three atoms of the acceptor group (denoting the lone-pair plane) and the water hydrogen atom were analyzed (Figure S1 – Supporting Information). The 3D plot in Table 3 (entries 1–6, 9–12) shows the distribution of the hydrogen atoms of water molecules for hydrogen-bond acceptors and how water is arranged along the LPVs. Interestingly, water hydrogen atoms in aliphatic and phenolic hydroxyls (entries 2, 3) are also positioned in between the two LPVs. Sulfoxide/sulfone and phosphine/phosphone (entries 7, 8) do not show a well-defined distribution because the S–O and P–O bond have only partial double-bond character.⁵⁷ In the latter two cases (entries 7, 8) water hydrogen atoms are distributed equally (i.e., without any spatial preference) forming a “corona”-like pattern around the acceptor oxygen atom. The deviation from the acceptor plane is generally below 30°. In our algorithm, the optimal positions for interaction with water are at the end points of LPVs and HEVs. Deviations from the LPV or HEV of up to 45° are allowed (Figure 3).

Table 4. Summary of the Crystallographic Data Search for Functional Groups Acting As Hydrogen Bond Donors (HBD)^a

HBA: Functional groups acting as hydrogen bond acceptors				
Entry	Query	Side view	Front view	Measurements
13	OH - Hydroxyl  Hits/Fragments: 1266/1814			d [3-4] (Å) angle [2-3-4] (°) Mean 1.80±0.21 162.70±15.0 Median 1.76 166.7 LQ/HQ 1.6/1.99 146.2/175.6
14	NH - Amino  Hits/Fragments: 1138/1599			d [3-4] (Å) angle [2-3-4] (°) Mean 2.0±0.23 156.7±16.6 Median 1.94 160.8 LQ/HQ 1.75/2.34 134.3/173.3
15	CH - Aromatic  Hits/Fragments: 758/1213			d [3-4] (Å) angle [2-3-4] (°) Mean 2.54±0.13 145.6±16.1 Median 2.56 145.8 LQ/HQ 2.35/2.69 124.3/167.6
16	CH - Aliphatic  Hits/Fragments: 305/433			d [3-4] (Å) angle [2-3-4] (°) Mean 2.59±0.10 146.9±15.0 Median 2.61 147.8 LQ/HQ 2.46/2.70 125.4/167.1

^aThe organization is identical to Table 3 with the only exception of water oxygen atoms, which are displayed with red spheres and the gray arrows, which refer to the HEVs.

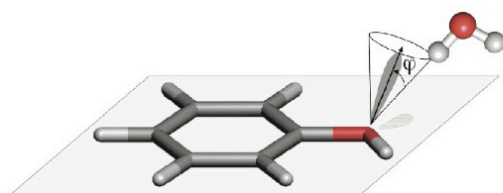


Figure 3. Definition of the allowed deviation (φ) from an ideal LPV (black arrow), in this case assumed at the lone-pair position. The same criteria are used for HEVs.

Despite similar studies,^{56–61} no analysis focusing exclusively on water molecules interacting with generic functional groups is available. Consistent with a previous study,⁵⁹ our analysis

revealed a clear sp^2 lone-pair preference of the carbonyl when interacting with water but not for the sp^3 lone pair for ether oxygen atom. A previous study showed hydrogen-bonding preference for heteroaromatic nitrogen acceptors compared to heteroaromatic oxygen acceptors.⁶⁰ Our results showed that water hydrogen atoms are less scattered around the lone pair of the nitrogen (entry 6) when compared to the distribution around the lone pairs of oxygen acceptor atom (entry 4).⁶⁰

Forty well-defined query hits (entry 12) for the interaction between water (as HBD) and aromatic rings suggest that even fragments generally considered as lipophilic can form favorable interactions with water molecules. In this case, the aromatic-ring atoms behave as HBAs and water molecules orient themselves to maximize the interactions of their hydrogen atoms with the π -electron cloud of the aromatic carbons, in a manner comparable with a hydrocarbon σ - π interaction.

H

dx.doi.org/10.1021/ci200150p | J. Chem. Inf. Model. XXXX, XXX, 000–000

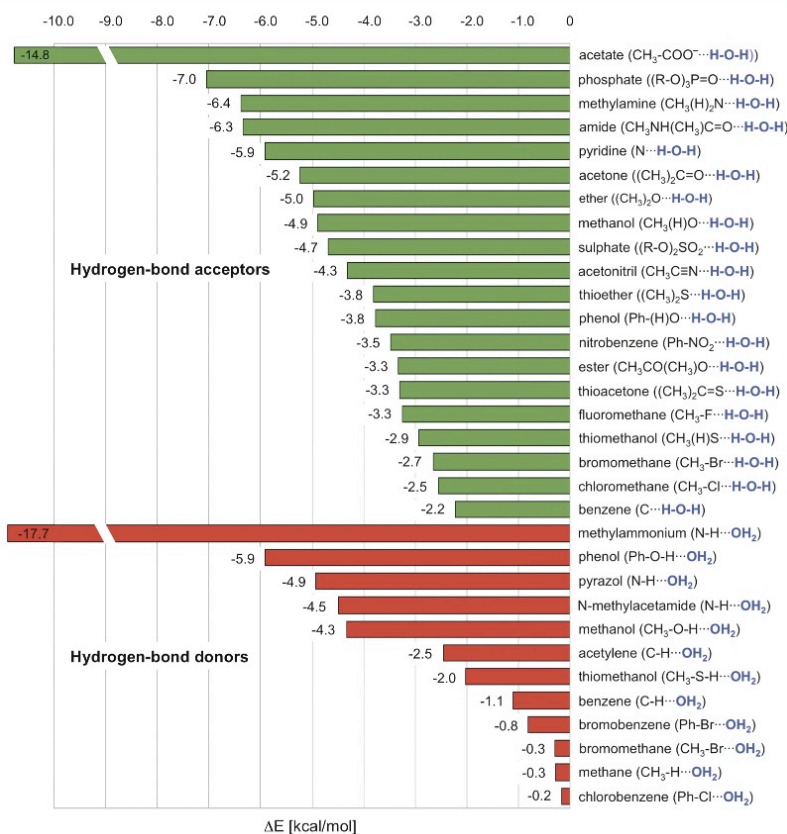


Figure 4. *Ab initio* calculated interaction energies for selected compounds interacting with water. Hydrogen-bond acceptors are depicted in green, hydrogen-bond donors in red.

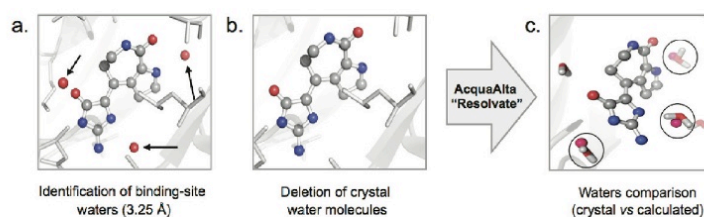


Figure 5. Algorithm validation. a) Identification of water molecules within 3.25 Å from each heavy atom of the ligand. b) Deletion of the water molecules. Application of the AcquaAlta solvation algorithm (gray arrow). c) Comparison between experimental and calculated waters (oxygen atom only).

Water oxygen atoms interacting with hydroxyl and amino HBDs (Table 4 – entries 13, 14) are distributed along the HEVs (OH and NH). Interestingly, in the case of the water oxygen interacting with apolar hydrogen atoms (Table 4 – entries 15, 16), the average median distance of 2.59 Å is lower than the sum of the van der Waals radii for hydrogen and oxygen (2.72 Å). This value, together with the range between lower and higher quantile

of the two queries (2.35–2.70 Å), shows that a weak interaction is established between the apolar aromatic and aliphatic hydrogen atoms and the oxygen of the water and that the water oxygen position is not a consequence of packing effects.⁵¹ [Quantiles: values marking equally sized consecutive subsets of an ordered sample of population (in our search analysis the quantile is set to 10).] The localization of the oxygen atoms of water shows no

Table 5. Results from the Validation Process^a

entry	PDB code	experimental binding site waters	match by AcquaAlta	identified bridging waters	match by AcquaAlta
1	2ayw	8	6 (75%)	3	3 (100%)
2	3dfr	23	12 (52.2%)	12	7 (58.3%)
3	1e2k	5	4 (80%)	2	2 (100%)
4	1ywn	2	1 (50%)	2	1 (50%)
5	1a8i	13	10 (77.0%)	9	6 (66.7%)
6	1xp0	1	1 (100%)	1	1 (100%)
7	1bjv	2	2 (100%)	1	1 (100%)
8	1gd1	15	11 (73.3%)	8	6 (75%)
9	1uy6	4	4 (100%)	4	4 (100%)
10	1xgj	2	2 (100%)	2	2 (100%)
11	2b53	3	2 (66.7%)	1	1 (100%)
12	1o86	6	5 (83.4%)	4	4 (100%)
13	1h1d	4	3 (75%)	2	2 (100%)
14	1hpx	4	2 (50%)	3	2 (66.7%)
15	1ndw	6	3 (60%)	2	2 (100%)
16	1u4d	5	3 (60%)	3	2 (66.7%)
17	1f0r	2	1 (50%)	1	1 (100%)
18	1kim	3	2 (66.7%)	3	2 (66.7%)
19	1xkk	2	2 (100%)	1	1 (100%)
20	1m17	1	1 (100%)	1	1 (100%)

^aListed are 20 PDB structures ordered by resolution (see Table 1) and identified by an entry number. Examined are (a) number of experimental binding site waters and corresponding match by AcquaAlta, (b) number of identified bridging waters and corresponding match by AcquaAlta. No direct correlation was found between water matching and B-factor range of the experimental waters.

preferred distribution. Based on atomic distances and lone pair deviations, we deduce that water molecules are indeed interacting with these fragments through electrostatic interactions and that their spatial orientation is mainly determined by the neighboring atoms.

Finally, one aim of the CSD search was to assess whether the small size and relatively high mobility of a water molecule has an impact on the geometry of hydrogen bonds. Our results (Table S2 – Supporting Information) are generally in agreement with values obtained from queries where water is represented only as a generic hydroxyl group. However, a detailed analysis of all the interactions shows that water acting as acceptor systematically forms shorter hydrogen bonds especially if interacting with strong donors (OH, NH) when compared to the generic hydroxyl query. In contrast, water acting as hydrogen bond donor forms longer hydrogen bonds than those found for the generic hydroxyl.

Thresholds and value ranges used in our approach are in agreement with the geometries obtained in the CSD search. In order to process protein structures with lower resolution and possibly less accurate geometries,^{62,63} AcquaAlta softens the corresponding geometric criteria during an iterative scan.

Functional-Group Hydration Propensity. Despite the availability of geometric criteria for describing hydrogen bonds (see Tables 3 and 4), rules to calculate associated energies (enthalpy, entropy) are scarce.^{64,65} In addition, the treatment of hydrogen-bond networks has only recently been addressed.^{66,67} Up to date, the NIST database⁶⁸ lists only a few experimentally determined parameters for strength of hydrogen bonds between water and small charged species like methylammonium and acetate.

Different hydration propensities of ligand functional groups and protein side chains have been elucidated through an extensive statistical analysis performed on 392 high-resolution ligand–protein

structures.³ A recent study⁶⁹ experimentally determined the relative basicity and therefore the strength of hydrogen-bond acceptors using Fourier transform infrared spectrometry (FTIR). When compared to NMR, UV, and IR techniques, this method allows to identify and to rank HBAs of polyfunctional bases. Due to the extreme diversity of the structures deposited in the CSD, hydrogen-bond frequencies and geometries can hardly be used to quantify their strengths for establishing quantitative structure–activity relationships. Nonetheless, the agreement between the abundance of certain hydrogen bonds in the CSD and the acceptor strength⁶⁹ can indicate the likelihood of hydrogen-bond formation.

In order to obtain relative interaction energies as well as geometric information, particularly valuable for interactions that are not well represented in the CSD, we performed a series of *ab initio* calculations on hydrated functional groups (Figure 4). Here, calculated hydrogen-bond interaction energies were used as a direct measure of a given functional group's propensity to hydration. A subsequent ranking based on those interaction energies inspired by a statistical analysis of 17 ligand atom types³ was used to fine-tune our own hydration-propensity scale for the AMBER* force-field atom types.⁷⁰

Oxygen atoms of carboxyls are most likely to be involved in hydrogen bond interactions. Their *ab initio* interaction energies compared to those of *sp*³ oxygen atoms of ethers and especially esters are substantially higher (ΔE ranges from -14.8 to -5.2 kcal/mol vs -5.0 to -3.3 kcal/mol). A similar trend for these two groups is observed when frequency of hits and number of fragments found in the CSD search are compared. Due to the low electronegativity of the phosphorus atom, the oxygen atom of a P=O group is also a strong acceptor ($\Delta E = -7.0$ kcal/mol), while the same atom in S=O or N=O groups forms comparably weaker hydrogen bonds (ΔE of -4.7 and -3.5 kcal/mol,

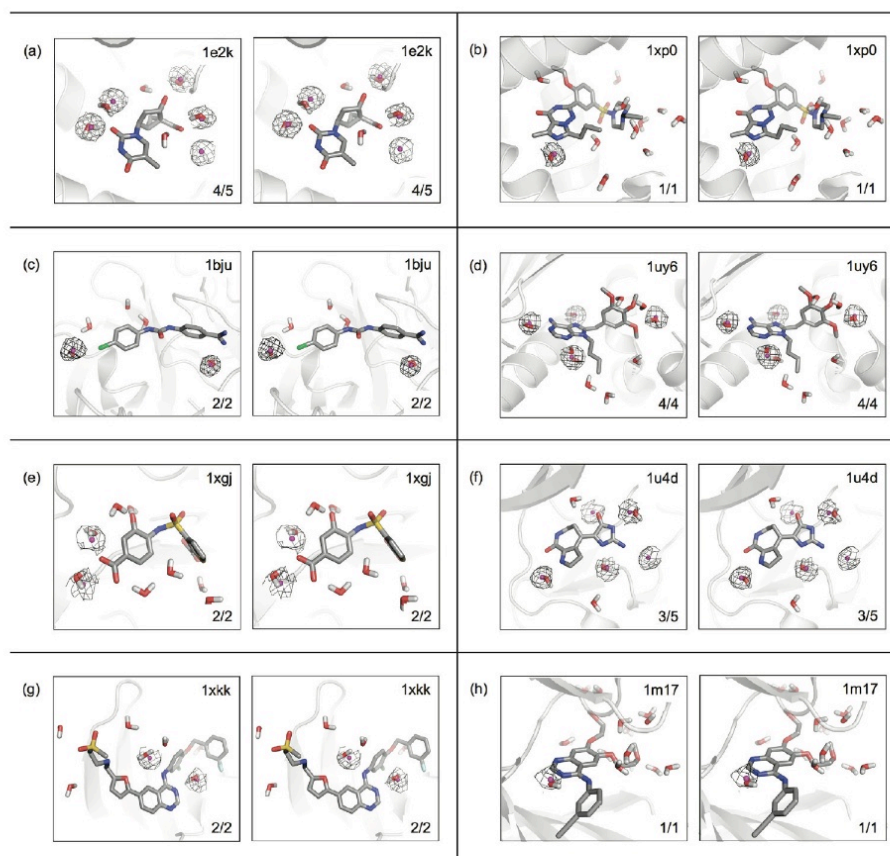


Figure 6. 3D representation (stereo view) of selected structures (PDB code in the top right corner) used in the validation. Ligand carbons are displayed in gray sticks, other ligand atoms are colored by atom type. Experimental waters are displayed as magenta spheres, water electron-density map as gray mesh and calculated waters in sticks representation. The match between experimental and calculated waters is given in the bottom right corner.

respectively). A nitrogen sp^2 atom incorporated in pyridine rings (-5.9 kcal/mol) forms a slightly stronger hydrogen bond than the sp hybridized nitrogen atom in acetonitrile (-4.3 kcal/mol). A sulfur atom, regardless of its hybridization state, was found to be a weak hydrogen bond acceptor, comparable to halogen atoms, where the fluorine atom was the most willing hydrogen bond acceptor. In the ideal case, aromatic carbon atoms can form a well-defined $\sigma-\pi$ interaction with a water molecule, contributing as much as -2.2 kcal/mol, which is roughly 40% of a strong hydrogen bond (e.g., $O-H \cdots O_{\text{acetone}} = -5.2$ kcal/mol).

When acting as a hydrogen-bond acceptor, water forms the strongest interactions with charged nitrogen groups. Otherwise no substantial difference in interaction strength (ΔE) can be found between N-H and O-H groups, with both being frequent and potent hydrogen-bond donors. Thiols form only weak interactions with water oxygen ($\Delta E = -2.0$ kcal/mol), as a matter of fact, they are even weaker than the best interaction formed with some hydrocarbons (e.g., for acetylene, $\Delta E = -2.45$ kcal/mol). Our calculation further suggests that halogens

form only weak interactions with water. The strongest halogen donor \cdots water interaction in Figure 4 is the one of bromobenzene featuring a ΔE of only -0.8 kcal/mol. Although, some of the listed interactions are quite rare (e.g., esters,⁵⁶ sulfones, and sulfonamides⁵⁸), they are included in Figure 4 to cover fragments as found in some of the CSD structures.

Validation of AcquaAlta. The AcquaAlta algorithm was validated on a data set of suitable crystal structures retrieved from the PDB. The structures belong to different pharmacological and biological target families. The resolution of the selected 20 crystal structures (Table 1) ranges from 0.97 Å for the bovine β -trypsin (PDB code: 2ayw) to 2.60 Å for the epidermal growth factor receptor (PDB code: 1m17). A subset of 15 structures (75%) showed resolutions lower or equal to 2.0 Å.

The most probable protonation state of both ligand and protein at physiological pH (7.4) was calculated using Epik.⁵² Then, calculated hydrogen atoms (as they are not typically resolved in the experimental structure) were added to the ligand and the protein structures by using the Yeti software. In the

K

dx.doi.org/10.1021/ct200150p | J. Chem. Inf. Model. XXXX, XXX, 000–000

Table 6. Results from Automated Docking: The 13 Peptides Aligned to the Template KNK Are Listed along with the “Self-Fit” Experiment (Entry # 5)^a

#	peptide aligned to KNK	matching possible PP	rmsd PP	rmsd aligned to crystal	match with crystal	match with aligned pose
1	KAK	26/28	0.486	0.623	4/7 (57.1%)	4/7 (57.1%)
2	KEK	27/31	0.768	0.634	5/9 (55.6%)	6/9 (66.7%)
3	KWK	27/31	0.923	0.916	5/6 (83.3%)	3/6 (50%)
4	KDK	28/30	0.805	0.619	5/10 (50%)	4/10 (40%)
5	KNK	32/32	0.550	0.495	5/7 (71.4%)	3/7 (42.9%)
6	KMK	26/31	0.811	0.513	3/7 (42.9%)	3/7 (42.9%)
7	KHK	28/32	0.975	0.933	5/7 (71.4%)	2/7 (28.6%)
8	KPK	24/25	0.597	0.831	5/6 (83.3%)	5/6 (83.3%)
9	KSK	28/29	0.721	0.660	7/9 (77.8%)	6/9 (66.7%)
10	KYK	27/30	1.068	2.885	5/7 (71.4%)	3/7 (42.9%)
11	KQK	30/33	1.271	1.280	6/10 (60%)	5/10 (50%)
12	KLK	27/30	0.839	1.230	5/6 (83.3%)	4/6 (66.7%)
13	KRK	27/37	0.533	0.855	5/6 (83.3%)	2/6 (33.3%)
14	KVK	25/30	0.768	0.540	2/6 (33.3%)	4/6 (66.7%)

^aIncluded are (a) number of the matched pharmacophore centers (PP) out of possible pharmacophores for each tripeptide, (b) the *rmsd* of the pharmacophores between template and aligned tripeptides, (c) the *rmsd* between the aligned tripeptide and the corresponding tripeptide from the crystal structure, (d) match of calculated to experimental waters when tripeptide from the X-ray structure and from the docking was used.

validation, we identified all experimental waters located within 3.25 Å from any ligand heavy atom (Figure 5a, waters indicated with arrows). To allow for an unbiased procedure, the experimental water coordinates were deleted (Figure 5b), and in the main step, were recalculated by AcquaAlta, oriented and minimized using the Yeti software, featuring a directional force-field. In the last step (Figure 5c), the positions of calculated waters were compared to the experimental position.

The match between experimental and calculated waters is summarized in Table 5. A filter (filtering criteria listed in the Methods section) to discriminate between binding site waters and bridging waters was applied. The match refers to calculated waters compared with experimental waters for each structure. Two different matches are provided considering experimental binding site waters or identified bridging waters.

The comparison (Table 5 and Figure 6) shows a 76% match between experimental and calculated binding site waters, while the match for the identified bridging waters was 87.5%. Analysis of the subset of 15 crystal structures with resolutions equal to or lower than 2.0 Å provided a match of 76.2% for binding site waters and a better prediction of 87.8% for bridging waters, showing a small but not significant improvement of the results.

In 6 out of 20 structures (entries 2, 4, 5, 14, 16, 18), the match of the bridging waters is lower or equal to 66.7%. Two structures (entries 2 and 5) contain rather high numbers of waters in the binding site (23 and 13, respectively). In both structures, the high number of water molecules originates from the presence of cofactors, which are also included in the solvation process, as well as from the ligands being localized in solvent accessible areas. Despite the high resolution of entry 2 (1.70 Å), only a low match (52.2% and 58.3% for binding site and bridging waters, respectively) is obtained, a fact that can probably be attributed to the localization of both ligand and cofactor on the enzyme surface where water networks can easily be formed. The poor result obtained for this structure can be explained by the fact that water-network geometries are not implemented in our approach.

The poor results obtained for entries 4, 14, 16, and 18, despite the ligand not being localized on the protein surface, are related to the concept of the algorithm. AcquaAlta aims to generate waters engaging in polar interactions with the ligand, therefore bridging interaction to polar atoms of the protein. Waters that were not matched by our algorithm (e.g., W163 of entry 4, W566 and W607 of entry 14, W503 and W603 of the chain b of entry 16, and W498 of entry 18) are bridging hydrogen bonds between the protein atoms and have only some electrostatic interactions with ligand atoms. Most of such unmatched waters are positioned beyond geometrical limits implemented in the AcquaAlta program, and their positions are primarily influenced by interactions with the protein atoms or “second shell” waters (e.g., W504 interacting with W503 of entry 16).

Selected examples of the results obtained in the validation phase (Table 5) are depicted in Figure 6. In all eight cases the match of the experimental binding site waters is higher or equal to the 60%. The calculated waters both match the experimental oxygen positions and fit the experimental water electron-densities.

APPLICATION OF ACQUAALTA TO AUTOMATED DOCKING

In order to challenge the solvation algorithm in a molecular docking application, we chose the oligopeptide binding protein (OppA) as a target. Fourteen crystal structures of the OppA with resolutions lower than 2.0 Å are available in the PDB (Table 2). In these structures, OppA is complexed with different tripeptides of the type lysine-X-lysine (K-X-K), where X refers to a variable amino-acid residue.

Molecular docking was performed in two steps. First, the prealignment/docking tool Alignator⁵⁴ was employed to align the tripeptides. The underlying protocol aligns low-energy conformers of the docked molecule, obtained from conformational searching in aqueous solution, to a template molecule, based on pharmacophores matching. Solutions having unfavorable close contacts with the protein, which may require substantial induced fit for a proper accommodation, are discarded.

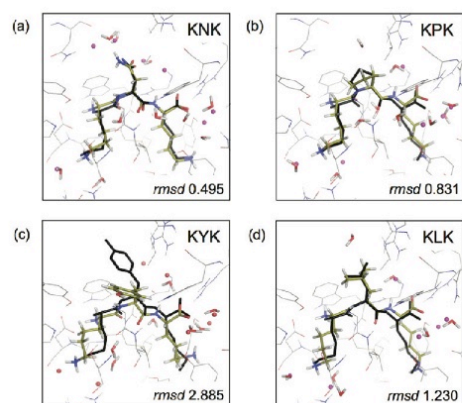


Figure 7. 3D representation of selected PDB structures (tripeptide sequence, upper right corner) from the automated docking. Crystal ligands are displayed in black sticks without oxygen atoms. Experimental waters are displayed as magenta spheres and calculated waters, based on the docked result, in sticks representation. In the bottom right corner the heavy-atoms *rmsd* of the oligopeptide (conformations between the crystal and the docked result) is given.

For this purpose, we selected the tripeptide KNK as template, which contains the highest number of diverse pharmacophore centers. The structural similarity of the docked tripeptides and the template is high, although the scaffold of KXX features a substantial number of rotatable bonds (>10). Moreover, for KXX motifs with X = D (aspartate) or X = E (glutamate), most of the low energy conformers have the two lysine side chains folded back to form an intramolecular stabilization. For this reason, the conformational search was set up with a rather large energy window (100 kcal/mol) and up to 12,000 conformers for each tripeptide were retained and aligned by Alignator. Accepting a wider variety of conformers increases the probability of identifying extended conformer geometries similar to those bound in the binding pocket.

The oligopeptides aligned to KNK displayed a high match of pharmacophore centers and in the case of the “self-fit” (i.e., conformers of KNK aligned to itself) all the pharmacophores between the docked tripeptide and the template were matched. The *rmsd* of the matched pharmacophores (Table 6) is lower than 1.0 Å in all the cases except of KYK (Figure 7c) where the tyrosine side chain adopts an alternative conformation compared to the crystal structure (χ_2 crystal = -61° , χ_2 docked = 58° , i.e. gauche⁻ vs gauche⁺) resulting in an apparent high *rmsd* value. The *rmsd* calculated on the heavy-atoms for aligned oligopeptides with the corresponding crystal conformation was always found to be lower than 1.0 Å, with the exception of KYK, KLK (Figure 7c,d), and KQK. High *rmsd* for KYK and KQK reflects different side chain orientation of the X amino acid. Alignments of tripeptides with long side chains (e.g., the two terminal lysine residues) are remarkably complex due to high number of possible combinations of their dihedral angles. Examples of aligned poses compared to the crystal structures, and experimental versus calculated water molecules, are shown in Figure 7.

After minimization, water molecules generated with AcquaAlta were compared with the experimental waters for both crystal

and Alignator-generated ligand poses, using the same distance criteria as in the validation phase. Neither geometry nor energy filters were applied. The average match for the 14 crystal poses is 66%, which is by 10% lower than the match obtained for the 20 complexes used in the validation. When the best *rmsd* pose from Alignator was used as input for the solvation algorithm, the resulting average match of water positions was 52.7%. A slight drop of accuracy can be expected, since in the adopted alignment protocol the conformers are aligned to the pharmacophore centers of a single oligopeptide template (KNK). In addition, as one can deduce from the *rmsd* scores, the matches are not perfect particularly for solvation sensitive hydrogen atoms in the terminal NH_3^+ group of the lysine side chain.

CONCLUSIONS

A correct simulation of ligands binding to proteins should account for structural and bridging solvent molecules. Based on a survey on 392 high-resolution crystal structure, an average of 4.6 water molecules are found at the ligand–protein interface, and three-quarters of them are involved in polar interactions with both ligand and protein.³

A rigorous analysis of small-molecule crystal structures from the CSD was performed to collect data on the geometry of interactions of water molecules when interacting with generic functional groups. *Ab initio* interaction energies between water and selected functional group representatives were calculated to construct an empiric hydration propensity scale. This information serves the basis for an algorithm (AcquaAlta), which solvates ligand–protein interfaces. AcquaAlta is based on the directionality of hydrogen bonds and can solvate both classical hydrogen bond moieties as well as halogen and hydrophobic functionalities of the ligand.

The approach was thoroughly validated on 20 X-ray structures with resolution ranging from 0.97 to 2.60 Å, checking the match of experimental and calculated water molecules. The match for binding site waters was 76%. The algorithm accuracy was not influenced substantially by the resolution of the crystal structures. When we applied geometry and energy filters to identify only the water molecules having polar contacts with both ligand and protein (i.e., bridging waters), the match rose to 87.5%. WATGEN,⁷¹ an algorithm to model water networks at protein–protein surfaces, yielded comparable results predicting 72% and 88% of water sites placed within 1.5 and 2.0 Å, respectively.

Subsequently, we combined our solvation approach with a pharmacophore-based alignment tool and applied it to 14 structures of OppA. This approach yielded poses within a *rmsd* of 1.0 Å from the crystal structures for 13 out of 14 complexes. The match, without any filtering criteria, between experimental and calculated water molecules was 66% and 53% when based on the crystal and docked poses, respectively. This drop in performance highlights the importance of a reasonable starting position to correctly reproduce water molecule positions. In a control test for the previously mentioned WATGEN algorithm,⁷¹ solvent molecules randomly placed at protein–protein interfaces were able to match only 22% and 40% of the experimental water sites within 1.5 and 2.0 Å.

In our approach, the solvated ligand–protein complex is not evaluated based on energy or on modification of entropy and enthalpy of the binding but rather on evaluation of the presence of hydrogen bond partners that a water molecule could bridge and thus bring to favorable interaction. The aim of this solvation algorithm is to produce crystal-like binding poses with optimally

arranged bridging waters, expecting that further refinement using force-field minimization routine would benefit and lead to a more accurate evaluation of thermodynamics of ligand binding. AcquaAlta is thought to be applied to docking studies where the presence or absence of a water molecule, calculated on each ligand conformation, can substantially affect pose scoring. On one side this could help obtaining reasonable binding modes; on the other one, the geometric criteria used do not give any information whether the water will enhance or worsen the binding. AcquaAlta is available on request through the Web site <http://www.modeling.unibas.ch/AcquaAlta>

■ ASSOCIATED CONTENT

● **Supporting Information.** Further details of the results from the CSD search, *ab initio* calculations, along with additional CSD searches. This material is available free of charge via the Internet at <http://pubs.acs.org>.

■ AUTHOR INFORMATION

Corresponding Author

*Fax: +41 61 267 1552. E-mail: martin.smiesko@unibas.ch. Corresponding author address: Department of Pharmaceutical Sciences, Pharmazentrum, University of Basel, Klingelbergstrasse 50, 4056 Basel, Switzerland.

■ ACKNOWLEDGMENT

Financial support by the Swiss National Foundation is gratefully acknowledged (grant number: 200021_119817).

■ ABBREVIATIONS:

CSD, Cambridge Structural Database; PDB, Protein Data Bank; HEV, Hydrogen-Extension Vector; LPV, Lone-Pair Vector; OppA, Oligopeptide binding protein A; HBA, Hydrogen Bond Acceptors; HBD, Hydrogen Bond Donors; QSAR, Quantitative structure–activity relationships

■ REFERENCES

- (1) Quiocho, F.; Wilson, D.; Vyas, N. Substrate specificity and affinity of a protein modulated by bound water molecules. *Nature* **1989**, *340* (6232), 404–407.
- (2) Poole, P. L.; Finney, J. L. Hydration-induced conformational and flexibility changes in lysozyme at low water content. *Int. J. Biol. Macromol.* **1983**, *5* (5), 308–310.
- (3) Lu, Y.; Wang, R.; Yang, C.; Wang, S. Analysis of ligand-bound water molecules in high-resolution crystal structures of protein–ligand complexes. *J. Chem. Inf. Model.* **2007**, *47* (2), 668–675.
- (4) ProteinDataBank. <http://www.pdb.org/> (accessed June 10, 2011).
- (5) Rodier, F.; Bahadur, R.; Chakrabarti, P.; Janin, J. Hydration of protein–protein interfaces. *Proteins* **2005**, *60* (1), 36–45.
- (6) Mancera, R. L. De novo ligand design with explicit water molecules: an application to bacterial neuraminidase. *J. Comput.-Aided Mol. Des.* **2002**, *16* (7), 479–499.
- (7) García-Sosa, A. T.; Mancera, R. L. The effect of a tightly bound water molecule on scaffold diversity in the computer-aided de novo ligand design of CDK2 inhibitors. *J. Mol. Model.* **2006**, *12* (4), 422–431.
- (8) Lloyd, D. G.; García-Sosa, A. T.; Alberts, I. L.; Todorov, N. P.; Mancera, R. L. The effect of tightly bound water molecules on the structural interpretation of ligand-derived pharmacophore models. *J. Comput.-Aided Mol. Des.* **2004**, *18* (2), 89–100.

- (9) Lam, P.; Jadhav, P.; Eyermann, C.; Hodge, C.; Ru, Y.; Bacheler, L.; Meek, J.; Otto, M.; Rayner, M.; Wong, Y. Rational design of potent, bioavailable, nonpeptide cyclic ureas as HIV protease inhibitors. *Science* **1994**, *263* (5145), 380–384.

- (10) Li, Z.; Lazaridis, T. Thermodynamic contributions of the ordered water molecule in HIV-1 protease. *J. Am. Chem. Soc.* **2003**, *125* (22), 6636–6637.

- (11) Ladbury, J. Just add water! The effect of water on the specificity of protein–ligand binding sites and its potential application to drug design. *Chem. Biol.* **1996**, *3* (12), 973–980.

- (12) Poornima, C.; Dean, P. Hydration in drug design. 1. Multiple hydrogen-bonding features of water molecules in mediating protein–ligand interactions. *J. Comput.-Aided Mol. Des.* **1995**, *9* (6), 500–512.

- (13) Poornima, C.; Dean, P. Hydration in drug design. 2. Influence of local site surface shape on water binding. *J. Comput.-Aided Mol. Des.* **1995**, *9* (6), 513–520.

- (14) Poornima, C.; Dean, P. Hydration in drug design. 3. Conserved water molecules at the ligand-binding sites of homologous proteins. *J. Comput.-Aided Mol. Des.* **1995**, *9* (6), 521–531.

- (15) Barillari, C.; Taylor, J.; Viner, R.; Essex, J. Classification of water molecules in protein binding sites. *J. Am. Chem. Soc.* **2007**, *129* (9), 2577–2587.

- (16) Fischer, S.; Verma, C. S. Binding of buried structural water increases the flexibility of proteins. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96* (17), 9613–9615.

- (17) Mancera, R. L. A new explicit hydration penalty score for ligand–protein interactions. *Chem. Phys. Lett.* **2004**, *399* (1–3), 271–275.

- (18) García-Sosa, A. T.; Firth-Clark, S.; Mancera, R. L. Including tightly-bound water molecules in de novo drug design. Exemplification through the in silico generation of poly(ADP-ribose)polymerase ligands. *J. Chem. Inf. Model.* **2005**, *45* (3), 624–633.

- (19) Li, Z.; Lazaridis, T. The effect of water displacement on binding thermodynamics: concanavalin A. *J. Phys. Chem. B* **2005**, *109* (1), 662–670.

- (20) Michel, J.; Tirado-Rives, J.; Jorgensen, W. Energetics of displacing water molecules from protein binding sites: consequences for ligand optimization. *J. Am. Chem. Soc.* **2009**, *131* (42), 15403–15411.

- (21) Mancera, R. L. Molecular modeling of hydration in drug design. *Curr. Opin. Drug Discovery Dev.* **2007**, *10* (3), 275–280.

- (22) García-Sosa, A. T.; Mancera, R. L. Free Energy Calculations of Mutations Involving a Tightly Bound Water Molecule and Ligand Substitutions in a Ligand–Protein Complex. *Mol. Inf.* **2010**, *29* (8–9), 589–600.

- (23) Michel, J.; Tirado-Rives, J.; Jorgensen, W. Prediction of the water content in protein binding sites. *J. Phys. Chem. B* **2009**, *113* (40), 13337–13346.

- (24) Abel, R.; Young, T.; Farid, R.; Berne, B.; Friesner, R. Role of the active-site solvent in the thermodynamics of factor Xa ligand binding. *J. Am. Chem. Soc.* **2008**, *130* (9), 2817–2831.

- (25) Pearlstein, R.; Hu, Q.; Zhou, J.; Yowe, D.; Levell, J.; Dale, B.; Kaushik, V.; Daniels, D.; Hanrahan, S.; Sherman, W.; Abel, R. New hypotheses about the structure–function of proprotein convertase subtilisin/kexin type 9: Analysis of the epidermal growth factor-like repeat A docking site using WaterMap. *Proteins* **2010**, *78* (12), 2571–2586.

- (26) Kortvelyesi, T.; Dennis, S.; Silberstein, M.; Brown, L. r.; Vajda, S. Algorithms for computational solvent mapping of proteins. *Proteins* **2003**, *51* (3), 340–351.

- (27) Ehrlich, L.; Reczko, M.; Bohr, H.; Wade, R. Prediction of protein hydration sites from sequence by modular neural networks. *Protein Eng.* **1998**, *11* (1), 11–19.

- (28) Amadasi, A.; Spyrikis, F.; Cozzini, P.; Abraham, D.; Kellogg, G.; Mozzarelli, A. Mapping the energetics of water–protein and water–ligand interactions with the “natural” HINT forcefield: predictive tools for characterizing the roles of water in biomolecules. *J. Mol. Biol.* **2006**, *358* (1), 289–309.

- (29) García-Sosa, A.; Mancera, R.; Dean, P. WaterScore: a novel method for distinguishing between bound and displaceable water molecules in the crystal structure of the binding site of protein–ligand complexes. *J. Mol. Model.* **2003**, *9* (3), 172–182.

- (30) Raymer, M.; Sanschagrín, P.; Punch, W.; Venkataraman, S.; Goodman, E.; Kuhn, L. Predicting conserved water-mediated and polar ligand interactions in proteins using a K-nearest-neighbors genetic algorithm. *J. Mol. Biol.* **1997**, *265* (4), 445–464.
- (31) Grant, J. A.; Pickup, B. T.; Nicholls, A. A smooth permittivity function for Poisson–Boltzmann solvation methods. *J. Comput. Chem.* **2001**, *22* (6), 608–640.
- (32) Verdonk, M.; Chessari, G.; Cole, J.; Hartshorn, M.; Murray, C.; Nissink, J.; Taylor, R.; Taylor, R. Modeling water molecules in protein-ligand docking using GOLD. *J. Med. Chem.* **2005**, *48* (20), 6504–6515.
- (33) Huang, N.; Shoichet, B. Exploiting ordered waters in molecular docking. *J. Med. Chem.* **2008**, *51* (16), 4862–4865.
- (34) Friesner, R.; Murphy, R.; Repasky, M.; Frye, L.; Greenwood, J.; Halgren, T.; Sanschagrín, P.; Mainz, D. Extra precision glide: docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. *J. Med. Chem.* **2006**, *49* (21), 6177–6196.
- (35) Corbeil, C.; Englebienne, P.; Moitessier, N. Docking ligands into flexible and solvated macromolecules. I. Development and validation of FITTED 1.0. *J. Chem. Inf. Model.* **2007**, *47* (2), 435–449.
- (36) de Graaf, C.; Oostenbrink, C.; Keizers, P.; van der Wijst, T.; Jongejans, A.; Vermeulen, N. Catalytic site prediction and virtual screening of cytochrome P450 2D6 substrates by consideration of water and rescoring in automated docking. *J. Med. Chem.* **2006**, *49* (8), 2417–2430.
- (37) Thilagavathi, R.; Mancera, R. Ligand-protein cross-docking with water molecules. *J. Chem. Inf. Model.* **2010**, *50* (3), 415–421.
- (38) Yang, J.; Chen, C. GEMDOCK: a generic evolutionary method for molecular docking. *Proteins* **2004**, *55* (2), 288–304.
- (39) Roberts, B. C.; Mancera, R. L. Ligand-protein docking with water molecules. *J. Chem. Inf. Model.* **2008**, *48* (2), 397–408.
- (40) Birch, L.; Murray, C.; Hartshorn, M.; Tickle, I.; Verdonk, M. Sensitivity of molecular docking to induced fit effects in influenza virus neuraminidase. *J. Comput.-Aided Mol. Des.* **2002**, *16* (12), 855–869.
- (41) Allen, F. The Cambridge Structural Database: a quarter of a million crystal structures and rising. *Acta Crystallogr., B* **2002**, *58* (Pt 3 Pt 1), 380–388.
- (42) Davies, T. G.; Hubbard, R. E.; Tame, J. R. Relating structure to thermodynamics: the crystal structures and binding affinity of eight OppA-peptide complexes. *Protein Sci.* **1999**, *8* (7), 1432–1444.
- (43) Tame, J. R.; Sleight, S. H.; Wilkinson, A. J.; Ladbury, J. E. The role of water in sequence-independent ligand binding by an oligopeptide transporter protein. *Nat. Struct. Biol.* **1996**, *3* (12), 998–1001.
- (44) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A.; Vreven, T.; Kudin, T. K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Adamo, C.; Jaramillo, C.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. In *Gaussian 2003*; J. A. Gaussian, Inc.: Pittsburgh, PA, 2003.
- (45) Moller, C.; Plesset, M. S. Note on an Approximation Treatment for Many-Electron Systems. *Phys. Rev.* **1934**, *46* (7), 618–622.
- (46) Head-Gordon, M.; Pople, J. A.; Frisch, M. J. MP2 energy evaluation by direct methods. *Chem. Phys. Lett.* **1988**, *153* (6), 503–506.
- (47) Simon, S.; Duran, M.; Dannenberg, J. J. How does basis set superposition error change the potential surfaces for hydrogen-bonded dimers? *J. Chem. Phys.* **1996**, *105*, 11024.
- (48) Boys, S. F.; Bernardi, F. The calculation of small molecular interactions by the differences of separate total energies. Some procedures with reduced errors. *Mol. Phys.* **1970**, *19*, 553.
- (49) Vedani, A.; Huhta, D. W. Algorithm for the systematic solvation of proteins based on the directionality of hydrogen bonds. *J. Am. Chem. Soc.* **1991**, *113* (15), 5860–5862.
- (50) Lu, Y.; Wang, Y.; Zhu, W. Nonbonding interactions of organic halogens in biological systems: implications for drug discovery and biomolecular design. *Phys. Chem. Chem. Phys.* **2010**, *12* (18), 4543–4551.
- (51) Taylor, R.; Kennard, O. Crystallographic Evidence for the Existence of C-H...O, C-H...N, and C-H...Cl Hydrogen Bonds. *J. Am. Chem. Soc.* **1982**, *104*, 5063–5070.
- (52) Schrödinger. <http://www.schrodinger.com/> (accessed June 10, 2011).
- (53) Vedani, A.; Huhta, D. W. A New Force-Field for Modeling Metalloproteins. *J. Am. Chem. Soc.* **1990**, *112* (12), 4759–4767.
- (54) Rossato, G.; Ernst, B.; Smiesko, M.; Spreafico, M.; Vedani, A. Probing Small-Molecule Binding to Cytochrome P450 2D6 and 2C9: An In Silico Protocol for Generating Toxicity Alerts. *ChemMedChem* **2010**, *5* (12), 2088–2101.
- (55) Klemperer, W. The Potential to Surprise. *Nature* **1993**, *362*, 698.
- (56) Murray-Rust, P.; Glusker, J. P. Directional hydrogen bonding to sp²- and sp³-hybridized oxygen atoms and its relevance to ligand-macromolecule interactions. *J. Am. Chem. Soc.* **1984**, *106* (4), 1018–1025.
- (57) Vedani, A.; Dunitz, J. D. Lone-pair directionality in hydrogen-bond potential functions for molecular mechanics calculations: the inhibition of human carbonic anhydrase II by sulfonamides. *J. Am. Chem. Soc.* **1985**, *107* (25), 7653–7658.
- (58) Bissantz, C.; Kuhn, B.; Stahl, M. A medicinal chemist's guide to molecular interactions. *J. Med. Chem.* **2010**, *53* (14), 5061–5084.
- (59) Lommerse, J. P. M.; Price, S. L.; Taylor, R. Hydrogen bonding of carbonyl, ether, and ester oxygen atoms with alkanol hydroxyl groups. *J. Comput. Chem.* **1997**, *18*, 757–774.
- (60) Nobeli, I.; Price, S. L.; Lommerse, J. P. M.; Taylor, R. Hydrogen bonding properties of oxygen and nitrogen acceptors in aromatic heterocycles. *J. Comput. Chem.* **1997**, *18*, 2060–2074.
- (61) Taylor, R.; Kennard, O.; Versichel, W. Geometry of the imino-carbonyl (N-H...O=C) hydrogen bond. I. Lone-pair directionality. *J. Am. Chem. Soc.* **1983**, *105* (18), 5761–5766.
- (62) DePristo, M.; de Bakker, P.; Blundell, T. Heterogeneity and inaccuracy in protein structures solved by X-ray crystallography. *Structure* **2004**, *12* (5), 831–838.
- (63) Henrik, C.; Görbitz, C. H.; Etter, M. C. Hydrogen bonds to carboxylate groups. The question of three-centre interactions. *J. Chem. Soc., Perkin Trans. 2* **1992**, 131–135.
- (64) Bohm, H. J.; Schneider, G. *Protein–Ligand Interactions from Molecular Recognition to Drug Design*; Wiley-VCH Verlag GmbH and Co.: Weinheim, Germany, 2003.
- (65) Baron, R.; Setny, P.; McCammon, J. A. Water in cavity-ligand recognition. *J. Am. Chem. Soc.* **2010**, *132* (34), 12091–12097.
- (66) Matthews, B. W.; Liu, L. A review about nothing: are apolar cavities in proteins really empty?. *Protein Sci.* **2009**, *18* (3), 494–502.
- (67) Plumridge, T. H.; Waigh, R. D. Water structure theory and some implications for drug design. *J. Pharm. Pharmacol.* **2002**, *54* (9), 1155–1179.
- (68) NIST Chemistry WebBook. <http://webbook.nist.gov/> (accessed June 10, 2011).
- (69) Laurence, C.; Brameld, K.; Graton, J.; Le Questel, J.; Renault, E. The pK(BHX) database: toward a better understanding of hydrogen-bond basicity for medicinal chemists. *J. Med. Chem.* **2009**, *52* (14), 4073–4086.
- (70) Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S.; Weiner, P. A New Force-Field for Molecular Mechanical Simulation of Nucleic-Acids and Proteins. *J. Am. Chem. Soc.* **1984**, *106* (3), 765–784.
- (71) Bui, H.; Schiewe, A.; Haworth, I. WATGEN: an algorithm for modeling water networks at protein-protein interfaces. *J. Comput. Chem.* **2007**, *28* (14), 2241–2251.

CURRICULUM VITAE

GIANLUCA ROSSATO



PERSONAL INFORMATION

Name Surname, Title **Gianluca Rossato, PharmD**
Date of birth 24/06/1983
E-mail gianluca.rossato@gmail.com
Address Hirzbodenweg 47, Basel – 4052, Switzerland
Residence via Conegliano 5, Oriago di Mira, Venice – 30034, Italy
Telephone +41 76 739 69 00
+39 340 365 23 80
Nationality Italian
Relationship status Single
Working permit (CH) Aufenthaltsbewilligung B – Swiss Working Permit (B) till 31.10.2013

CURRENT POSITION

Nov. 2008 – Present.
(PhD defense: July 2011)

PhD in Pharmaceutical Sciences (Molecular Modeling and Drug Design).
PhD advisor: Prof. Angelo Vedani
Institute of Molecular Pharmacy, Department of Pharmaceutical Sciences
University of Basel (*Switzerland*)

Thesis title: Molecular simulations on proteins of biomedical interest:

- A) In silico solvation of ligand–protein complexes
- B) QSAR for Cytochrome P450 2D6 and 2C9
- C) Structure–based design at the Myelin Associated Glycoprotein (MAG)

Drug Discovery area:

- Computer-Aided Drug Design
- Molecular Modeling and Simulation
- Toxicology Prediction

FORMER POSITIONS / EDUCATION

2008 **Researcher** – University of Helsinki, (*Finland*)
Center for Drug Research (CDR)
Computational Drug Discovery group – Group Leader: Dr. Henri Xhaard

2002 – 2008 **M.Sc. Pharmacy** – University of Padova (*Italy*)
Experimental thesis:
Structural Bioinformatic Laboratory (SBL) – Åbo Akademi University, Turku (*Finland*)
Supervisors: Prof. Mark S. Johnson (Åbo Akademi University)
Prof. Stefano Moro (University of Padova)

Thesis title: “Comparative Modeling of dopamine, norepinephrine and serotonin transporters in complex with tricyclic antidepressants.”

2002 High school certificate – Mestre – Venice – Italy

<p>PROFESSIONAL SKILLS</p>	<p>Molecular Modeling and Simulation: ligand and structure-based drug design</p> <p>Software knowledge:</p> <ul style="list-style-type: none"> • Operating systems: Unix (different distributions), Windows, Macintosh OsX • Modeling tools: <ul style="list-style-type: none"> – Schrodinger suite: Macromodel, Glide, Prime, Desmond – Molecular visualization tools: Pymol, VMD, Swiss PDB viewer – Others: AutoDock, Quasar, Raptor, Sybyl, Yeti • Programming: C++ (one year experience), bash scripting
<p>SCIENTIFIC PUBLICATIONS</p>	<ol style="list-style-type: none"> 1. <u>G. Rossato</u>, A. Vedani, M. Smiesko. AcquaAlta: an algorithm for the dynamic solvation of protein–ligand complexes. <i>J. Chem. Inf. Model.</i> 2011, (Accepted) 2. <u>G. Rossato</u>, B. Ernst, M. Smiesko, M. Spreafico, A. Vedani. Probing small-molecule binding to Cytochrome P450 2D6 and 2C9: An in silico protocol for generating toxicity alerts. <i>ChemMedChem.</i> 2010, 5, 2088-2011 3. S. Mesch, D. Moser, D. S. Strasser, A. Kelm, B. Cutting, <u>G. Rossato</u>, A. Vedani, H. Koliwer-Brandl, M. Wittwer, S. Rabbani, O. Schwardt, S. Kelm, B. Ernst, Low Molecular Weight Antagonists of the Myelin-Associated Glycoprotein: Synthesis, Docking, and Biological Evaluation. <i>J. Med. Chem.</i> 2010, 53(4), 1597-615 4. O. Schwardt, H. Koliwer-Brandl, R. Zimmerli, S. Mesch, <u>G. Rossato</u>, M. Spreafico, A. Vedani, S. Kelm, B. Ernst. Design, synthesis, biological evaluation, and modeling of a non-carbohydrate antagonist of the myelin-associated glycoprotein. <i>Bioorg. Med. Chem.</i> 2010, 18, 7239-7251 5. M. Spreafico, M. Smiesko, O. Peristera, <u>G. Rossato</u>, A. Vedani. Probing small-molecule binding to the Liver-X receptor: A mixed-model QSAR study. <i>Mol. Inf.</i> 2010, 1, 27–36 6. B. Ernst, O. Schwardt, S. Mesch, M. Wittwer, <u>G. Rossato</u>, A. Vedani. From the Ganglioside GQ1b to Glycomimetic Antagonists of the Myelin-Associated Glycoprotein (MAG). <i>CHIMIA International Journal for Chemistry</i>, 2010, 64(1), 17-22.
<p>AWARDS</p>	<ul style="list-style-type: none"> • Poster Award – Annual Research Meeting 2010 - Pharmazentrum - University of Basel <i>Poster title:</i> Multi-dimensional QSAR at cytochrome P450 2D6; A pharmacophore-based, flexible docking approach. • “Student Grant”, (2009) Banca di Monastier e del Sile, Italy • “Young Talented Researchers Grant”, (one-year scholarship, 2008). Research Foundation of the University of Helsinki.

<p>DRUG DEVELOPMENT and CLINICAL TRIAL EDUCATION</p>	<ul style="list-style-type: none"> • <u>Key Issues in Drug Discovery & Development (KIDD)</u> (13 Jan – 04 Feb, 2011) Graduate Course – University of Basel Drug-Development modules: Clinical Phases I – III; GCP; Registration; Clinical Phase IV (drug regulatory affairs, pharmacoconomics and reimbursement, phase IV trials and drug safety, patenting and licensing). • <u>Essential in Drug Development & Clinical Trials</u> (24 Nov – 23 Dec, 2010) Swiss Tropical and Public Health Institute – Basel Topics treated: Drug development organization, regulations, GCP & Authorities, Monitoring & Auditing, Study design and Trial statistics, Research Ethics. Case Study: Study design of a comparative trial for antibiotics used in nosocomial infections. • <u>GCP and Regulatory Requirements</u> (5–8 May, 2010) Clinical Trial Unit, CTU – Basel University Professional in Clinical Trial Practice and Management Topics treated: ICH-GCP guideline E6; Ethics in clinical trials; regulatory and legal frameworks; clinical trials directive; clinical trial authorization; national implementation; pediatric regulation.
<p>FURTHER COMPUTER SKILLS</p>	<p>MS Office suite: Word, Excel, Powerpoint</p> <p>Graphic / Web Design: Adobe Photoshop, Macromedia Flash</p>
<p>LANGUAGE SKILLS</p>	<p>English: Fluent TOEFL 99/120 (20-09-2008), further improved by 3 years of daily work in English.</p> <p>Italian: Native speaker</p> <p>German: Beginner</p>
<p>ADDITIONAL INFORMATION</p>	<ul style="list-style-type: none"> • Speaker Annual Research Meeting 2011– Pharmazentrum, Basel • Scientific writing trainee – MediWrite GmbH, Basel (Mar – May, 2010) • Tutoring experience of pharmacy students during M.Sc thesis. (Jan – Jun 2009) • Habilitation as pharmacist, pharmacist trainee (2007); pizza-maker trainee (2006) • Hobbies: traveling, darkroom films/photos development, mountain biking, writing • European driving license (B)