Analysis, Design and *in silico* Evaluation

of E-selectin Antagonists

Inauguraldissertation

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

von

Michele Giovanni Antonio Porro

aus Chiasso (Schweiz)

Basel, Juni 2006

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät

Auf Antrag von:

Prof. Dr. B. Ernst, Institut für Molekulare Pharmazie, Universität Basel,

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

Prof. Dr. M. Hamburger, Institut für Pharmazeutische Biologie, Universität Basel

Basel, den 14. Februar 2006

Prof. Dr. Hans-Jakob Wirz

Dekan

Declaration

"I declare, that I wrote this thesis *Analysis, Design and in silico Evaluation of E-selectin Antagonists* with the help indicated and only handed it in to the faculty of science of the University of Basel and to no other faculty and no other university."

Michele Porro, Basel, the 23rd of January 2006

Acknowledgements

Il ringraziamento più grande lo devo ai miei genitori per avermi permesso di studiare e di raggiungere oggi questo importante traguardo. Il loro sostegno é sempre stato incondizionato e prezioso, anche nei momenti più difficili. GRAZIE !

Ringrazio molto anche mio fratello Marco, con cui ho condiviso a lungo l'appartamento di Basilea e tante "avventure". In lui ho sempre trovato qualcuno con cui poter discutere. Negli ultimi tempi poi, il suo trasferimento in Lussemburgo mi ha permesso di avere un altro posto dove rifugiarmi durante i week-ends. Lì, ho conosciuto anche Anna che ringrazio per le ottime cene e soprattutto per la pasta di olive.

Sul piano scientifico, non posso far a meno di ringraziare il Professor Beat Ernst per avermi permesso di svolgere il mio lavoro di dottorato presso l'instituto di farmacia molecolare dell'università di Basilea. Il suo supporto e la libertà d'azione che mi ha lasciato, mi hanno dato la possibilità di svilupparmi come ricercatore e di vivere esperienze uniche come i congressi di Shanghai e Siena.

Un ringraziamento particolare va al Professor Angelo Vedani. Non l'avessi incontrato sette anni fà in una piccolissima stanza del vecchio instituto di farmacia in Tötengässlein, probabilmente oggi non avrei in mano una tesi in bioinformatica. Inoltre, senza i suoi consigli, i suoi programmi (comprese tutte le modifiche che ha dovuto apportare a causa mia…) e soprattutto senza le sue correzioni questo lavoro sarebbe stato molto più difficile. Grazie mille, Angelo!

Ringrazio il dottor Markus Lill per avermi introdotto all' uso di *AMBER* e per le interessanti discussioni.

Un ringraziamento va anche a Berna, al mio collega Samuel Schmid, per aver scritto diversi scripts e il programma *SAMOA* che mi sono stati così utili nell'analisi dei miei dati. Il suo supporto "tecnico" e le molte discussioni telefoniche sono impagabili.

Un grosso grazie lo devo a Monika Schmid e Eva Oetliker. I loro lavori di diploma sono parte integrante di questa tesi.

Anche se arrivata da poco, ringrazio Morena Spreafico per il buon umore e il sostegno negli ultimi mesi.

Un grazie lo devo anche a tutti i miei colleghi dell'instituto di farmacia molecolare.

In particolare vorrei ringraziare l' "E-selectin team" (Dr. S. Marti, Dr. C. Müller, Dr. Z. Dragic, Dr. O. Schwardt, Dr. S. Rabbani, Dr. B. Cutting, A. Titz, L. Tschopp, D. Schwizer, Y. Jing, B. Wagner e il Prof. B. Ernst) con cui ho potuto lavorare in maniera piacevole e interessante.

I miei "vicini di laboratorio": con Steven, Tamara, Daniel Strasser, Karin e Daniel Ricklin il lavoro in laboratorio non é mai stato un peso.

Abstract

E-selectin, is member of a family of cell-adhesion proteins, which plays a crucial role in many physiological processes and diseases [1], and in particular, in the early phases of the inflammatory response. Its role is to promote the tethering and the rolling of leukocytes along the endothelial surface [2]. These steps are then followed by integrin-mediated firm adhesion and final transendothelial migration. Therefore, control of the leukocyte-endothelial cell adhesion process may be useful in cases, where excessive recruitment of leukocytes can contribute to acute or chronic diseases such as stroke, reperfusion injury, psoriasis or rheumatoid arthritis [3].

In this work, efforts to develop *in silico*-based protocols to study the interaction between E-selectin and its ligands, are presented. Hence, different protocols had to be developed and validated. In particular, a new procedure for the analysis of the conformational preferences of E-selectin antagonists was established and the results compared to those obtained with the MC(JBW)/SD approach, which had already demonstrated its validity in the past [161,168]. Thus, the comparison between the two protocols permitted to recognize a different conformational preference of the two methods for the orientation of the sialic acid moiety of sLe^x (**3**) (torsions Φ_3 and Ψ_3, *Figure* A), which reflects the contrasting opinions existing for the conformation adopted by sLe^x (**3**) in solution [150–168]. A more detailed analysis revealed that probably both approaches deliver only a partially correct view and that in reality, in solution, sLe^x (**3**) exists as a mixture of low energy conformers and not as supposed to date [150–154,161–163] as a population of a single conformer.
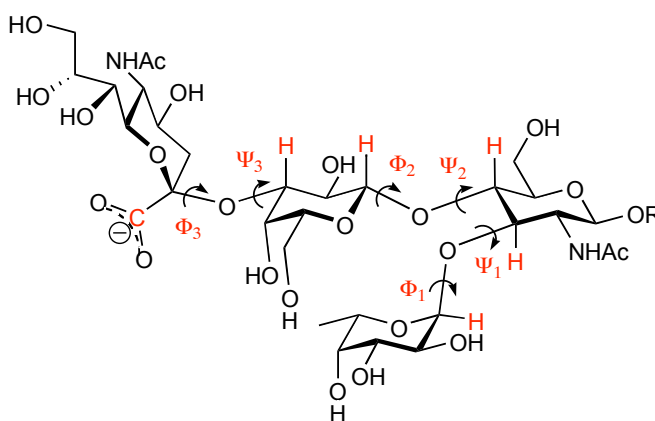
Figure A: sLe^x (**3**) and the Φ, Ψ convention for the definition of the glycosidic torsions.

In addition, a docking routine was established and the impact of different partial-charge methods and of explicit solvation on the binding mode studied.

MD simulations enabled to gain an insight into the dynamical character of the protein-ligand interactions. In particular, the observations done in an atomic-force microscopy study [350], describing the interactions between the carboxylic group of sLe^x and Arg97, and between the 3– and 4–hydroxyls of fucose and the calcium ion, as the two main energy barriers for the dissociation process of the protein-ligand complex, found confirmation in our MD-investigations. Thus, these two contacts always lasted longer than any other in the MD simulation.

QSAR-models with *Quasar* [270–272,351] and *Raptor* [315,316,335] were successfully derived and will permit a semi-quantitative *in silico* estimation of the binding affinity for the ligands that will be designed in the future.

Finally, the developed protocols and models were applied for the development of new E-selectin antagonists. Unfortunately, to date, only few biological data is available to evaluate our design strategies. However, the impact of the ligand's pre-organization on the binding affinity could be established at least for the Le^x-core of sLe^x (**3**). Hence, the importance of the exo-anomeric effect, of the steric compression, and of the hydrophobic interaction between the methyl group of fucose and the β-face of galactose was clearly demonstrated.

Abbreviations

Ac	Acetyl
AM1-BCC	Austin Model 1 with bond charge corrections
ASGR	Asialoglycoprotein receptor
ax	Axial
Bn	Benzyl
BSA	Bovine serum albumine
Bz	Benzoyl
cAMP	Cyclic adenosine mono phosphate
CD18	Integrin β 2, Cell differentiation antigen 18
CD34	Cell differentiation antigen 34, sialomucin
cf.	Compare for example
CM	Charge Model
CoMFA	Comparative molecular field analysis
CoMSIA	Comparative molecular similarity indices analysis
CR	Complement regulatory-like domains
CRD	Carbohydrate recognition domain
CS	Conformational searches
ECM	Extracellular matrix
e.g.	For example
EGF	Epidermal growth factor
ELISA	Enzyme-linked immunosorbent assay
ESL-1	E-selectin ligand 1
ESP	Electrostatic potential
FEP	Free energy perturbation
Fuc	Fucose
Gal	Galactose
GB/SA	Generalized-born surface-area
Glc*N*Ac	*N*-Acetylglucosamine
Glu	Glucose
Gly-CAM-1	Glycosylated cell adhesion molecule-1
GOLPE	General optimal PLS equation
GvHD	Graft vs host disease
H	Hydrogen
HEV	High endothelial venules
HUVEC	Human umbilical vein endothelium cells
JBW	Jumping between wells
K_D	Dissociation constant
k_{off}	Dissociation rate costant

k_{on}	Association rate costant
IC_{50}	Inhibitory concentration 50%
rIC_{50}	Relative IC_{50}
ICAM-1	Intercellular cell adhesion molecule 1
IFN$-\gamma$	Interferon$-\gamma$
IL-1	Interleukine-1
IL-1$-\beta$	Interleukine 1$-\beta$
IL-3	Interleukine-3
LAD	Leukocyte adhesion deficiency
LDL	Low density lipoprotein
Le^X	Lewis^X
LoF	Lack of fit
LPS	Lipopolysaccharide
MadCAM-1	Mucosal vascular addressin cell adhesion molecule
Man	Mannose
MBP	Mannose binding protein
MC	Monte-Carlo
MD	Molecular dynamics
Me	Methyl
MM	Molecular Mechanics
MM/JBW	MacroModel/ Jumping between Wells
MMPs	Matrix Metalloproteinases
MNDO	Modified neglect of differential overlaps method
Neu, NeuAc	Neuraminic acid = Sialic acid
NF$-\kappa$B	Nuclear factor of kappa B
NMR	Nuclear magnetic resonance
NO	Nitric oxide
nOe	Nuclear Overhauser effect
OH	Hydroxyl
p^2	Predictive r^2
Ph	Phenyl
PLS	Partial least square
PSGL-1	P-selectin glycoprotein ligand 1
q^2	Cross-validated r^2
QM	Quanto mechanics
QSAR	Quantitative structure-activity relationship
r^2	Correlation coefficient
rms	Root mean square
SAR	Structure-activity relationship
SD	Stochastic dynamics
Sia	Sialic acid = Neuraminic acid
sLe^a	Sialyl Lewis^a

sLe^X	Sialyl Lewis^X
SPR	Surface plasmon resonance
SUMM	Systematic multiple minimum search
Temp	Temperature
TNF$-\alpha$	Tissue necrosis factor α
VCAM-1	Endothelial vascular cell-adhesion molecule-1

Table of Contents

1 Introduction

1.1 The choice of the target-protein

E-selectin, an adhesion molecule presently under investigation at the Institute of Molecular Pharmacy of the University of Basel, is part of a protein family, which plays a crucial role in many physiological processes and diseases [1]. More specifically, selectins are a family of carbohydrate-binding proteins expressed at the site of inflammation in response to mediators of inflammation liberated by the injured tissue. Their key role, early in the inflammatory cascade, is to promote the tethering and the rolling of leukocytes along the endothelial surface [2]. These steps are then followed by integrin-mediated firm adhesion and final transendothelial migration.

Therefore, control of the leukocyte-endothelial cell adhesion process may be useful in cases, where excess recruitment of leukocytes can contribute to acute or chronic diseases such as stroke, reperfusion injury, psoriasis or rheumatoid arthritis [3]. In addition, it has recently been suggested that cancer may exploit the adhesion process after entering the bloodstream to metastasize [4,5].

Thus, due to its implication in pharmacologically important processes, E-selectin has emerged as an interesting therapeutic target for the pharmaceutical industry as well as for academic research.

1.2 Inflammation

Inflammation is a complex stereotypical reaction of the organism in response to the damage of cells and of vascularized tissues. By studying the details of the processes associated with inflammation, a close relationship to the immune response has been revealed.

Already *Aulus Cornelius Celsus* (ca. 25 B.C.–A.D. 50) has described the basic symptoms of inflammation – *rubor* (redness), *tumour* (swelling), *calor* (heat), *dolor* (pain) and *functio lesa* (deranged function). Such signs are the typical consequence of the extravasation of leukocytes from the plasma to the site of inflammation [6]. Early investigations considered the inflammatory reaction as a primary host defense system [7]. However, it has been showed that an excessive response can lead to debilitating diseases e.g. arthritis and gout or even to death, as in the case of an anaphylactic shock [8].

1.2.1 The inflammatory cascade

In a healthy person, the invasion of a tissue by pathogens or a tissue injury usually triggers the release of inflammatory mediators like chemokines or platelet-activating factors [9,10]. This first stimulus initiates a complex response that involves the activation and directed migration of leukocytes (neutrophils, monocytes and eosinophils) from the venous system to the sites of damage by a complex series of steps, referred to as the inflammatory cascade (*Figure* 1).
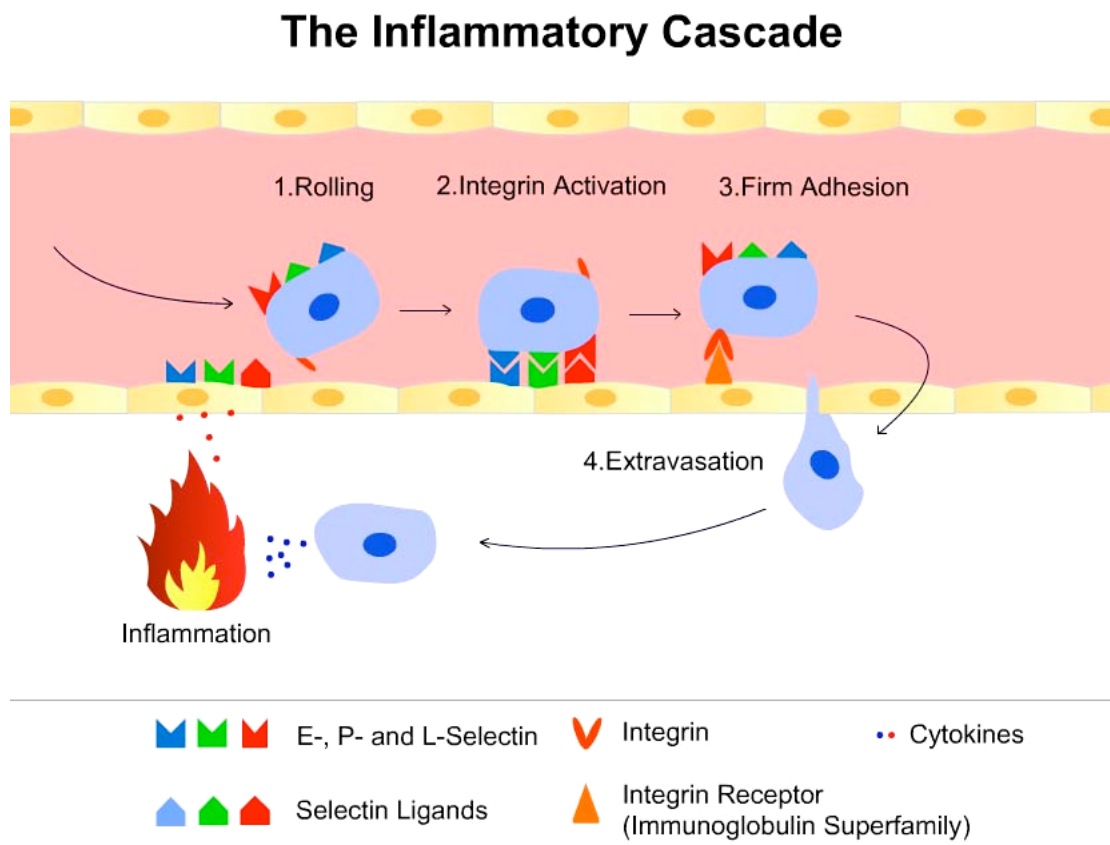
Figure 1: The inflammatory cascade.

The inflammatory cascade can be summarized in five major steps:

Step 1: *Stimulus*

After tissue injury or pathogens invasion, the immune system responds with the release of a variety of inflammatory mediators. Such substances (e.g. TNF-α,IL-1 or LPS) stimulate endothelium cells to transiently express E- and P-selectin [11,12]. P-selectin, which is stored in the α-granules of platelets and in the *Waibel-Palade bodies* of endothelial cells, is rapidly transported to the cell surface and expressed within seconds to minutes after the stimulation occurred by pro-inflammatory substances such as histamine or thrombin [13,14]. The highest expression level on the cell surface is reached after some five to ten minutes. After 30-60 minutes, the protein is internalized again by endocytosis. In addition to this fast exposure of P-selectin, a second mechanisms involving the activation of E-selectin exists. E-selectin, in contrast to P-selectin, has to be synthesized *de novo*. The production of E-selectin is induced by TNF-α, IL-1 or LPS [15,16]. The highest expression levels on the cell surface are reached after three to four hours and basal levels can still be detected after 16-24h [17]. As observed for P-selectin, also E-selectin starts to be internalized by endocytosis few hours later it has reached its highest expression level on the cell surface [18].

Step 2: *Tethering and Rolling*

Next, the presentation of E- and P-selectin at the surface of endothelial cells leads to the interaction of these proteins with their natural ligands (ESL-1, PSGL-1,…) present on the surface of the leukocytes. The fast association and dissociation process (cf. *Chapter* 1.4.3) of leukocytes to endothelial cells produces the well-studied phenomena of tethering (attachment) and rolling of white blood cells along the vessel wall [19,20] (*Figure* 2).
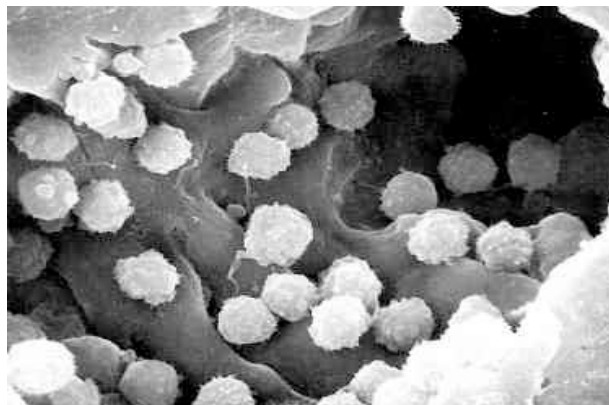
Figure 2: Leukocytes rolling on the surface of a blood vessel.

L-selectin, which is constitutively expressed on leukocytes, also plays an important role in this step: first, it contains carbohydrate structures that can be recognized by E-selectin [21], second, its interaction with PSGL-1 [22] of already adhering leukocytes [23]. This last mechanism permits to expand the number of layers of leukocytes attracted to the site of inflammation.

The processes of tethering and rolling are prerequisites for the adhesion step and are therefore essential to the inflammatory process. Hence, selectins became an important therapeutic target (cf. *Chapter* 1.3.2).

Step 3: *Secondary activation*

The rolling of leukocytes along the vessel wall enables the interaction of the vascular endothelium with cytokines and leukocytes-activating molecules that trigger the activation and upregulation of leukocytes integrins [24,25]. Integrins, which are also important drug targets for inflammatory diseases [26–30], form a second class of leukocyte adhesion molecules that play an essential role in the inflammatory cascade, in particular in the fourth step: the adhesion step.

Step 4: *Adhesion*

Integrins, in fact, are needed to immobilize the leukocytes onto the endothelium surface. The immobilization or tight adhesion is reached through the interaction of the integrins with their endothelial ligands VCAM-1, ICAM-1, and MAdCAM-1 [24,25].

Step 5: *Transendothelial migration*

When the leukocytes are adhered to the endothelium, they can transmigrate (extravasation) to the site of inflammation (*Figure* 3). This process is probably facilitated by extracellular proteases, such as matrix metalloproteinases (MMPs).
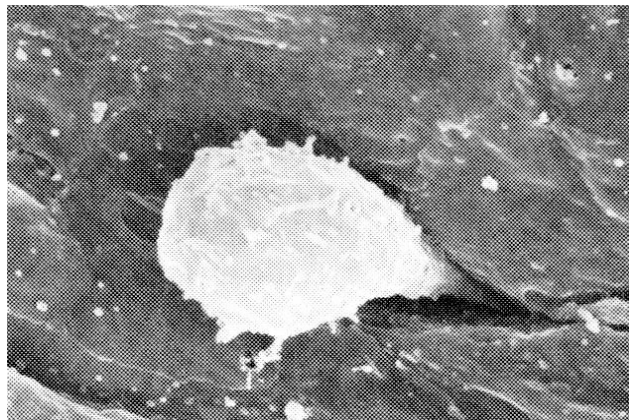
Figure 3: A leukocyte transmigrating into the site of inflammation.

1.3 The pathological role of selectins

1.3.1 Observations in mice

Numerous studies have confirmed the involvement of all three selectins in the inflammatory cascade [31–34]. Experiments with gene-deficient mice (knockout mice) permitted elucidate the physiological role of the selectins.

By studying L-selectin knockout mice, significant reduction in lymphocyte homing [35,36] and an insufficient immune response was observed [37–39]. P-selectin knockout mice showed reduced neutrophil transmigration into the inflamed peritoneum [40,41]. In contrast, E-selectin knockout mice didn't present any abnormalities in the inflammatory response [42,43]. However severe defects were observed in those mice, when P-selectin was blocked by P-selectin antibodies [44] or also mutated. Particularly, the double mutant mice demonstrated an increased susceptibility to bacterial infection, a strong reduction of the number of rolling leukocytes, and the complete absence of netrophil transmigration in first four hours following an inflammatory stimulus [43,45].

1.3.2 Selectin and human diseases

Excessive leukocyte accumulation in inflamed tissues is the basis for a pathological inflammatory reaction, angiogenesis and tumor metastasis. In diseases such as atherosclerosis, asthma, organ rejection after transplantation, hemorrhagic shock, thrombosis, rheumatoid arthritis, atopic dermatitis, psoriasis, diabetes-caused microangiopathies, or myocardial and cerebral ischemia, a strong deregulation of the selectin expression and function has been observed [46–48].

In the case of atherosclerosis, accumulation of LDL (low density lipoprotein) in blood vessel walls triggers an inflammatory fibroid reaction. As a consequence, E- and P-selectins are expressed on the surface of the endothelial cells thereby facilitating the invasion of the inflamed tissue by monocytes. The correlation between the high expression levels of selectins and the development of a large number of atherosclerosis plaques was demonstrated in animal models [49–52].

The role of selectins in myocardial or cerebral ischemia/reperfusion has been studied extensively [53]. When myocardial ischemia occurs, a loss of endothelium-derived nitrogen monoxide and a rapid burst of oxygen-derived radicals arise. Those provoke an upregulated expression of P-selectin on the cells surface and thereby a strong accumulation of neutrophils. This accumulation in the already damaged tissue induces a vascular dysfunction and further severe damages to the heart muscle cells. The prevention of reperfusion injury emerged therefore as an important therapeutic goal. *In vivo* it was possible

to show that the administration of P-selectin antibodies [54] or sLe^x-related oligosaccharides [55] had a protective effect.

One of most prominent autoimmune diseases, rheumatoid arthritis, is caused by a chronic reaction of the immune system. Pathological tissue observations demonstrated an inflammation of the synovial membrane of the joints, characterized by a massive infiltration of leukocytes. Being pressed against the surrounding bones by the presence of the leukocytes, the synovial membrane permanently releases inflammatory mediators, leading to a chronic state. Although the exact molecular mechanism of the rheumatoid arthritis is not fully understood, the role of E-selectin in the pathogenesis and its accumulation in rheumatoid tissue has been recognized. Soluble E- and P-selectin serum levels are therefore used as the molecular markers of active inflammation in rheumatoid arthritis [56–59].

Selectins are also involved in asthma [48]. E-selectin, in particular, together with the integrin receptors ICAM-1 and VCAM-1 [60] mediates the recruitment of eosinophil granulocytes in the lungs.

In addition to that, increased expression of selectins has been related to the rejection of human liver [61], cardial [62,63], bone marrow and renal transplants [64], implying therefore a role of these molecules in this complex process. By patients, who developed the so-called "graft vs host disease" (GvHD), a multiorgan disease caused by immune response against the donated bone marrow, increased levels of E-selectin and VCAM-1 were detected [65,66].

Recent data have given an insight to the fact that inflammation is a critical component of tumor progression [67]. Through the years, it has become clear that the tumor microenvironment, which is largely populated by inflammatory cells, actively participates to tumor proliferation, survival and migration. As a matter of fact, it is now known that tumor cells can adhere to the endothelium exploiting the mechanism of leukocyte homing [68–72]. Hence, in the case of colon and breast carcinoma, a correlation between the metastatic behavior of the tumor and the expression of sialidated E-selectin ligands on the tumor cell surface has been established [73–76]. Recently, it was even showed that the expression level of sLe^a and a poor prognosis of survey of colon cancer clearly correlate [77].

Selectins are also related to a very rare genetic disease called "type 2 leukocyte adhesion deficiency (LAD-2)" [78,79]. Patients affected from LAD-2 lack of fucosylated glucoconjugates and can therefore not express functional selectin ligands. Clinically, they show mental retardation, short stature and recurrent bacterial infection with high leukocyte counts. This medical picture, underscores the fundamental role of selectins in the initial recruitment of leukocytes into the site of inflammation or infection. In contrast, LAD-1 patients lack of functional integrin β_2-chains (CD18), essential for neutrophil extravasation and suffer from life-threatening infections [80].

In summary, selectins-related diseases may be classified into five groups:

- Acute allergy-related diseases: such as stroke, reperfusion injury during myocardial or cerebral ischemia, organ transplantation, and traumatic or hemorrhagic shock

- Chronic allergic diseases: such as asthma, psoriasis, dermatitis and rheumatoid arthritis

- Cardiovascular diseases: such as arteriosclerosis and peripheral vascular disease

- Cancer: Angiogenesis and cancer metastasis

- Metabolic diseases, such as LAD-2

1.4 The selectin family: a class of versatile proteins

1.4.1 E-, P-, and L-selectin

Lectins are carbohydrate-binding proteins and are typically divided in four groups [81]:

- C-type lectins, which bind one or more calcium-ions in the carbohydrate binding site (therefore C-type, C = Calcium)

- S-lectins or galectins containing free thiol groups (therefore S, S = Sulphur)

- P-lectins, recognizing phoshorylated mannose residues (therefore P, P = Phosphor) and

- Other lectins not fitting in one of the above categories.

E-selectin, like the high homologous L- and P-selectin, is a C-type lectin. All the three selectins present a calcium-binding domain in the carbohydrate recognition domain (CRD). Other examples for C-type lectins are the asialo-glycoprotein receptor (ASGR) [82] and the mannose binding protein (MBP) [83,84]. The selectin family itself consists of just the three proteins named before: E-selectin, P-selectin and L-selectin.

The prefixes (E-, P-, and L-) of the three selectins indicate the cell types where the molecules have been identified first: E-selectin on activated **e**ndothelial cells, L-selectin on **l**eukocytes and P-selectin in the *Waibel-Palade bodies* of endothelial cells and in α-granules of **p**latelets [85]. Selectins belong to type I membrane glycoproteins and share a common structural motif including a N-terminal C-type lectin domain (CRD), an epidermal growth factor-like (EGF-like) domain, a variable number of complement regulatory-like repeats, referred to as "consensus repeat" or "complement regulatory-like" (CR) domain, a transmembrane domain, and a short cytoplasmic tail [86] (*Figure* 4). The

sequence identity among the different selectin species amounts to 52% in the CRD, 47% in the EGF domain, and to 35% in the CR domain.
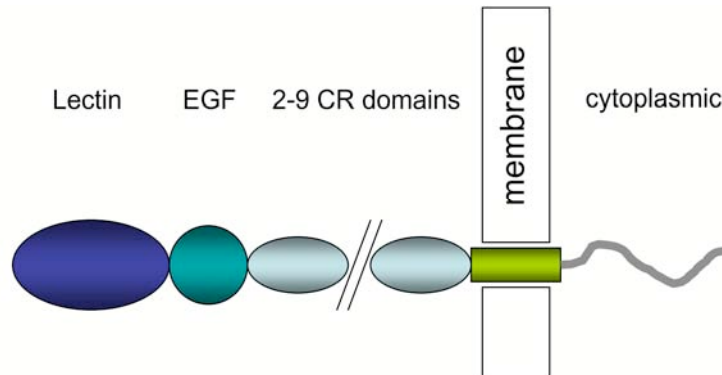
Figure 4: Domains organization of the selectins.

The C-terminal portion of the protein is located intracellular and its functional role, which is still not fully elucidated, is supposed to be related to signal transduction [87]. The cytoplasmatic tail is rather short, and comprises as few as 17 amino acids in L-selectin, 32 in E-selectin, and 35 in P-selectin. After a short transmembrane domain, the CR domains are located. These short repetitive elements are about 60 amino acids long, contain three disulfides bridges, and its numbers differs throughout the selectin family (*Figure* 4). Humans E-selectin contains six, P-selectin nine, and L-selectin only two CR domains. Among other species, the number of CRs varies from four to eight. The functional role of the CRs is still not fully elucidated but several experiments suggest that the removal of several CRs impair the efficiency of the selectins to support leukocyte rolling [88–91]. These findings, along with studies that showed that only the CRD and the EGF-like domain are needed for specific ligand binding [91], suggest that the role of the CRs is to guarantee sufficient distance between the plasma membrane and the CRD. Following the CRs is the EGF-like domain. It contains six cysteins, located at equivalent positions in the "EGF-repeats" of several proteins. Although the carbohydrate binding-site was identified within the CRD for all selectins [92], the EGF-domain appears to be necessary for ligand binding. The EGF-domain is supposed to either stabilize the conformation of the CRD or to directly interact with the ligand [89,93–99]. The N-terminal domain, also referred to as CRD, is composed of approximately 120 amino acids and presents the typical features of a lectin domain of C-type animal lectin [100]. It bears the carbohydrate binding-site [92], conformationally stabilized by a calcium ion.

1.4.2 The natural glycoprotein ligands of the selectins and their carbohydrate epitopes

Selectins are carbohydrate-binding proteins. Their natural ligands are therefore glycoproteins or glycolipids, which presents glycan structures as a binding motif. It is generally accepted that to achieve binding to the target protein, selectin-ligands have to present as an epitope either the trisaccharides Lewis[x] (Le[x]) (**1**) or Lewis[a] (**2**) or their sialylated derivatives sialyl Lewis[x] (sLe[x]) (**3**) or sialyl Lewis[a] (sLe[a]) (**4**) (*Figure* 5) [101–103]. In some circumstances, additional sulfation of the saccharides is needed to promote binding.
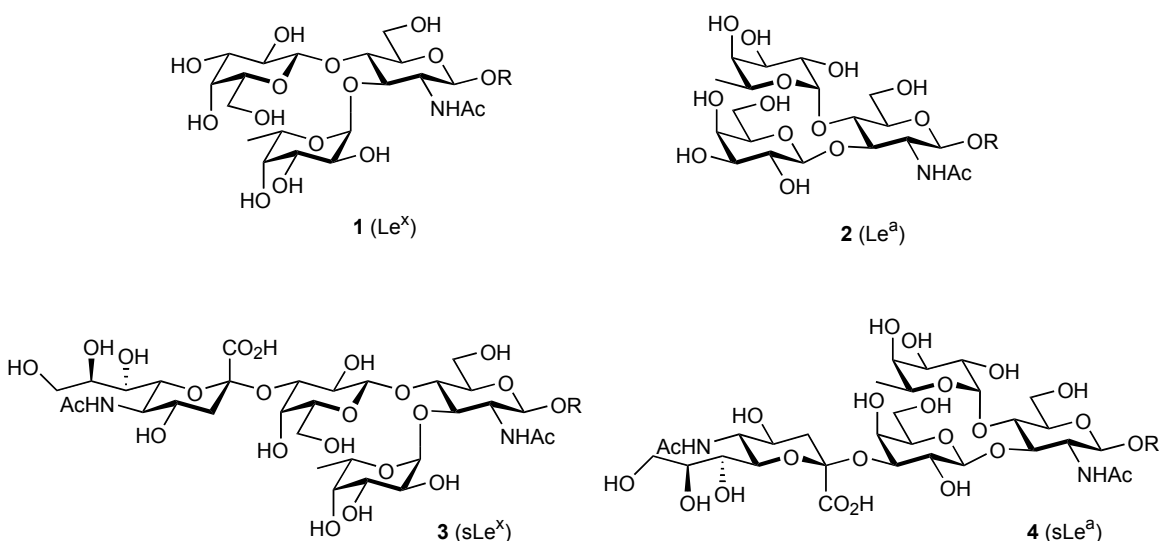
Figure 5: Selectin ligands contain the common carbohydrate epitope sLe[x] (**3**), which was shown to interact with all three selectins, albeit with different affinities.

Soluble recombinant forms and IgG fusion protein-constructs of selectins were used to identify their natural ligands. As of today, five glycoproteins binding to L-selectin were discovered: MAdCAM-1 [104,105], Spg200 [106], Gly-CAM-1 [107], CD34 [108] and the podocalyxin-like protein [109]. All of them are expressed by high endothelial venules in lymph node tissue. MAdCAM-1, which contains both a mucin- and an immunoglobulin-like domain, usually binds the lymphocyte integrin $\alpha_4\beta_7$. However, one of its subpopulations can also be recognized by L-selectin.

Gly-CAM-1 and CD34 are sialomucins presenting large clusters of sialic acid-rich *O*-linked carbohydrate side chains that seem to be essential for L-selectin binding. Both proteins are also expressed in tissues other than the lymph nodes

but there they lack the required carbohydrate modifications to promote binding. Gly-CAM-1 itself is a secretory protein, which is usually stored in cytoplasmic granula [110,111]. Its posttranslational modifications have been intensively studied and revealed that sulfation, sialidation and fucosylation are essential to promote the binding to L-selectin [107,112,113].

In 1993 a glycoprotein, referred to as PSGL-1, was identified as a P-selectin ligand [114]. PSGL-1 is a 250kDa homodimeric protein linked by two disulfide bridges. Like the L-selectin ligands, PSGL-1 also is a sialomucin with a high degree of *O*-linked glycan modifications. Beside the need of sialidation and fucosylation to bind P-selectin and probably also L-selectin, PSGL-1 has to be sulfated at two of the three N-terminal tyrosine residues (Tyr46 and one of the tyrosines Tyr48 or Tyr51) [115–118].

In the screening for E-selectin ligands, a glycoprotein referred to as ESL-1 was identified. ESL-1 was isolated from myeloid cells and neutrophils of mouse origin and characterized as a 150kDa glycoprotein [119]. In contrast to the glycoproteins discussed above, ESL-1 requires *N*-linked glycan modifications for binding E-selectin. Interesting ESL-1 binds only E- and not P-selectin [120]. Three other glycans binding to E-selectin were further identified [121]. All three contain the sialyl di-Le^x motif on the β-D-Glc*N*Ac-(1–4)-α-D-Man-(1–3)-branch of the tetra-antennary N-glycans. The specificity of these ligands was proved in an assay with an affinity column charged with soluble E-selectin [122]. E-selectin recognizes also PSGL-1 [120,123–127], even when the N-terminal tyrosine of the ligand are not sulfated [123,126]. Finally, it has been demonstrated that E-selectin binds the carbohydrates present on L-selectin on human neutrophils but not those located on lymphocytes [21,127].

The interactions between the three selectins and their binding partners are summarized in *Figure* 6 [129].
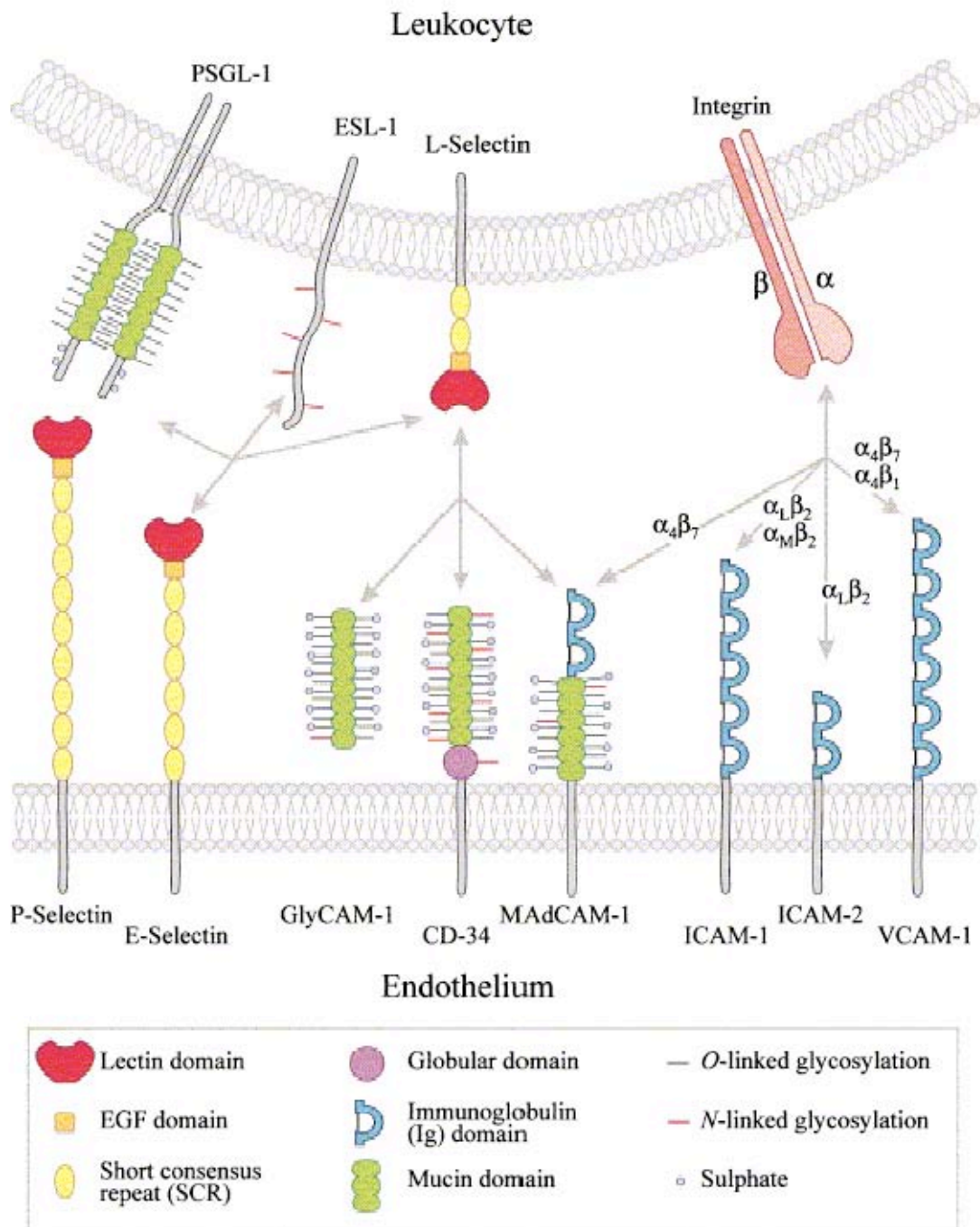
Figure 6: Selectins, integrins and their binding partners. The depicted selectin ligands are those, which have been identified by affinity isolation with the respective selectin as affinity probe [129].

1.4.3 Affinities and kinetics of the selectin-ligand interactions

The interactions between the selectins and their ligands are essential to slow down the leukocytes streaming in the blood vessels and initiate their rolling. This step is fundamental to further permit the leukocytes to adhere to the endothelial surface and then extravasate towards the site of inflammation. In the "rolling phase" the binding between the selectins and their partners must still be reversible. Indeed, the formation of the bond has to be very fast otherwise the leukocytes could not be slowed down. On the other hand, the dissociation reaction has to be slow enough to facilitate adhesion but also fast enough to ensure cellular integrity. Consequently, this special kind of cell-cell interaction requires low affinity (K_D), fast association (k_{on}) and fast dissociation (k_{off}) rates. In addition, this mechanism allows the resistance to the laminar shear stress caused by the blood stream. The interaction between the selectins and their ligands fulfills all these requirements [25,130,131]. A wealth of investigations has shown that the selectins bind synthetic oligosaccharide such as sLe^x (**3**) or sLe^a (**4**) with low affinity (K_D ranged between 0.1 to 5 mM) [132–136] and rapid kinetic. In a recent study [137], performed using surface plasmon resonance (SPR), the binding constance (K_D) as well as the k_{on} and k_{off} could be determined for the interaction between mouse E-selectin and ESL-1. This data and other coming from earlier observations [138,139] are presented in *Table* 1.

Table 1: Comparison of affinities and kinetics of selectin-ligand interactions.

Interaction	Species	Temp (°C)	K_D (µM)	k_{on} (M^{-1} s^{-1})	k_{off} (s^{-1})	Ref.
E-selectin / ESL-1	Mouse	37	62	7.4×10^4	4.6	[137]
		25	56	4.8×10^4	2.7	[137]
L-selectin / GlyCAM-1	Mouse	25	108	$> 1 \times 10^5$	> 10	[139]
P-selectin / PSGL-1	Human	25	0.32	4.4×10^5	1.4	[138]

In general, it can be stated that k_{on}-values of selectin ligands lie in the same range as they generally do for protein-carbohydrate interactions [140] and are marginally slower than typical protein-protein values (10^5 to 10^6 M^{-1} s^{-1}).

A closer look at the k_{off} values reveals how fast the whole process is. Hence, the lifetime of a selectin-ligand bond is of few seconds.

It is important to note that there is no significant change in the kinetic values when the temperature is changed, which implies that the binding is mostly driven by favorable entropic contributions.

As will be discussed later, the development of effective E-selectin antagonists has to take into account the associated kinetics particularity. Thus, with the aim to increase the affinity (K_D) to low μM or even better nM, k_{on} as well as k_{off} values have to be improved. Based on the aforementioned entropic considerations, a first approach to increase affinity would be to constrain the ligand in its bioactive conformation and/or to simplify the ligand by retaining only the pharmacophoric groups. As a consequence of these modifications, unfavorable contributions of entropy and desolvation to the ligand-binding energy could possibly be reduced. Thus, the k_{on} value would decrease, thereby positively influencing the binding affinity. On the other hand, k_{off} values could be increased by adding groups (e.g. hydrophobic substituents) that would favorably interact with the protein, resulting in higher enthalpic contributions.

1.5 Structure-activity relationship of sLe^x binding to E-selectin

The development of low-molecular weight, high affinity ligands for E-selectin, requires a comprehensive understanding of the 3D structure of sLe^x, of E-selectin itself, and of the ligand-protein complex. During the last 20 years, NMR-spectroscopy, X-ray diffraction studies, molecular simulations as well as binding-affinity studies with different E-selectin antagonists, have revealed a detailed picture of the structure-activity relationship. However, some details are still controversially today and some points remain unclear.

1.5.1 Pharmacophores

The identification of residues critical for sLe^x binding to the selectins illustrates an interesting aspect of contemporary organic synthesis. In fact, by modifying all functional groups of sLe^x in a systematic fashion, it was possible to determine, those essential for binding. Such groups are referred to as pharmacophores. As it can be seen from *Figure* 7, the pharmacophore of sLe^x is defined by:

1) the three hydroxyl groups of fucose,

2) the 4– and the 6–hydroxyl group of galactose,

3) the carboxylate group of neuraminic acid.
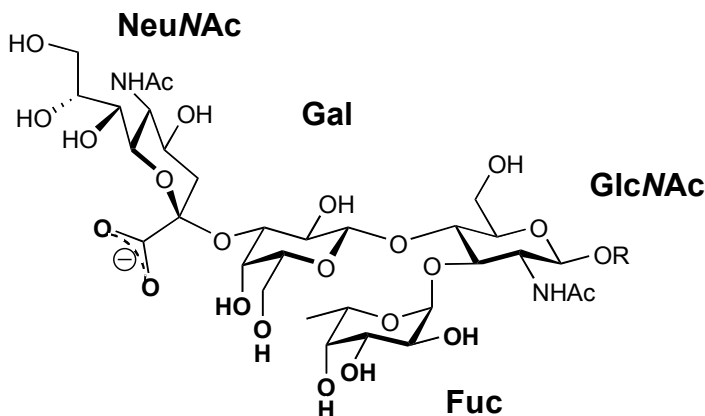
Figure 7: Structure/function map of sLe^x (**3**). Pharmacophores for binding to E-selectin are highlighted.

The role of the hydroxyl groups of fucose was first determined by Gaeta *et al.* [141] and by Hasegawa *et al.* [142]. They replaced the hydroxyl groups with hydrogen atoms and, as a result, binding to E-selectin was no longer observed. Correctly, they hypothesized that, as observed in a similar protein (MBP-A) [84], fucose was responsible for the binding to a calcium ion. In the case of P-selectin, however, only the 3-hydroxyl group turned out to be crucial for binding.

To investigate the role of the hydroxyl groups of galactose, deoxy- and fluoro-derivatives of sLe^x were synthesized [143]. Substitution of the 4– or the 6–hydroxyl groups showed a decrease in binding affinity, implying that those group are essential for binding.

Investigations of the functional groups of the Neu*N*Ac moiety (the glycerol side chain, the 4-hydroxyl group, the amide residue and the carboxylate) lead to the

conclusion that only the carboxylate plays a determinant role in the recognition of sLe[x] by E-selectin [142,144,145].

Several studies [146–148] also discuss the contribution of the GlcNAc moiety to the binding affinity. They all agree on the fact that, GlcNAc doesn't directly contribute to the protein-ligand interaction, but serves as a spacer unit to arrange the crucial functional groups of fucose and galactose in the required 3D-position and orientation.

1.5.2 Conformational studies on sLe[x]

1.5.2.1 The conformation of sLe[x] in solution

Early work in the selectin field was devoted to the study of the conformation of sLe[x] in water. NMR studies with labeled and unlabeled compounds, in combination with molecular-dynamics simulation, were performed. Initially, three independent studies [149] suggested the presence of a single conformation of sLe[x] in water. However, this finding was contradicted later [150,151]. A summary of the data presents in the literature can be found in *Table* 2. The convention for the definition of the glycosidic torsions, which will be used throughout the thesis, is presented in *Figure* 8.
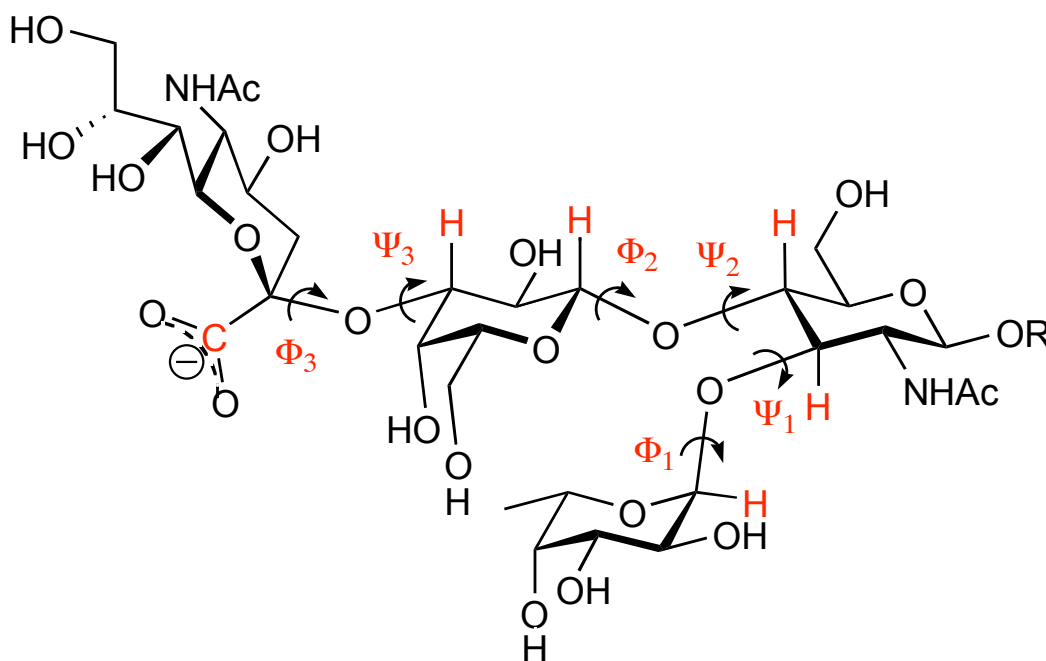
Figure 8: Φ, Ψ convention for the definition of the glycosidic torsions.

Table 2: Summary of the data presents in the literature for the solution conformation of sLe^x.

Φ_1 (°)	Ψ_1 (°)	Φ_2 (°)	Ψ_2 (°)	Φ_3 (°)	Ψ_3 (°)	%-populated	Ref.
48	25	55	7	163	−57	0% (A)	[152]
				−170	−8	100% (B)	[152]
				−79	7	0% (C)	[152]
				68	−20	0% (D)	[152]
23	30	48	15	167	−63	0%	[152]
24	30	48	16	−171	−6	0%	[152]
48	22	50	15	−95	−45	0%	[150]
				−70	5	100%	
				−160	−20	0%	
48	24	46	18	-	-	-	[153]
−23	15	46	18	-	-	-	[153]
65	40	65	15	−95	−60	80%	[154]
				−70	0	10%	[154]
				−160	−25	10%	[154]

It is worthwhile to note that there is a general agreement between the different studies on the conformation adopted by the so-called core structure formed by the Le^x-trisaccharide consisting of fucose, GlcNAc, and galactose, whereas substantial disagreement reigns regarding the conformation adopted around the NeuNAc-Gal linkage. This disagreement arises from whether or not a nuclear Overhauser effect (nOe) between the H–3 of Gal and the H–3(ax) of NeuNAc is observed (*Figure* 9). In fact, studies [152,155] observing this nOe suggest that the conformation around the NeuNAc-Gal linkage is described by the dihedral

angles {$\Phi_3 = -180° \pm 10°$, $\Psi_3 = 0° \pm 10°$}. Those [150,156] not observing the very nOe, in contrast suggest {$\Phi_3 = -70° \pm 10°$, $\Psi_3 = 0° \pm 10°$}.
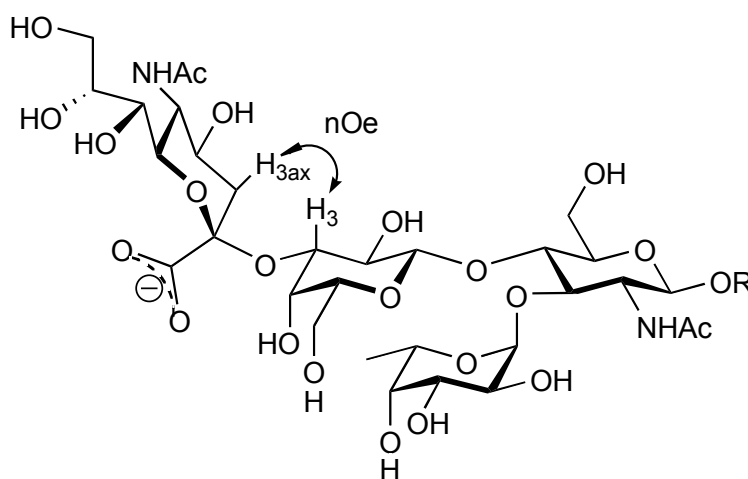
Figure 9: The strongly debated nOe.

It is, however, known [150,157] that the glycosidic bond between Neu*N*Ac and Gal is highly flexible. Hence, this observation could suggest the presence of multiple conformations of sLe[x] in water. As mentioned above, also theoretical methods, such as molecular dynamics simulations (MD), were used to determine the solution conformation of sLe[x]. MD bases on the fact, that a long enough simulation should produce a Boltzmann-weighted ensemble statistically representing the conformations of sLe[x] in water. Moreover, through the Karplus equations (*Figure* 10), especially parameterized for carbohydrates by Tvaroska [158] (*Equation* 1) and by Bose [159] (*Equation* 2), it is possible, to calculate $^3J_{C-H}$ and $^3J_{C-C}$ coupling constants and compare them with the experimental data [160]. An agreement between theoretical and experimental coupling constants, as obtain for example by Kolb and Ernst in their study [161] suggests that the structures sampled during a MD correctly display the distribution of conformations reigning in water. As a consequence, theoretical methods became a powerful tool for the design of E-selectin antagonists (cf. *Chapters* 1.5.3, 4.7).
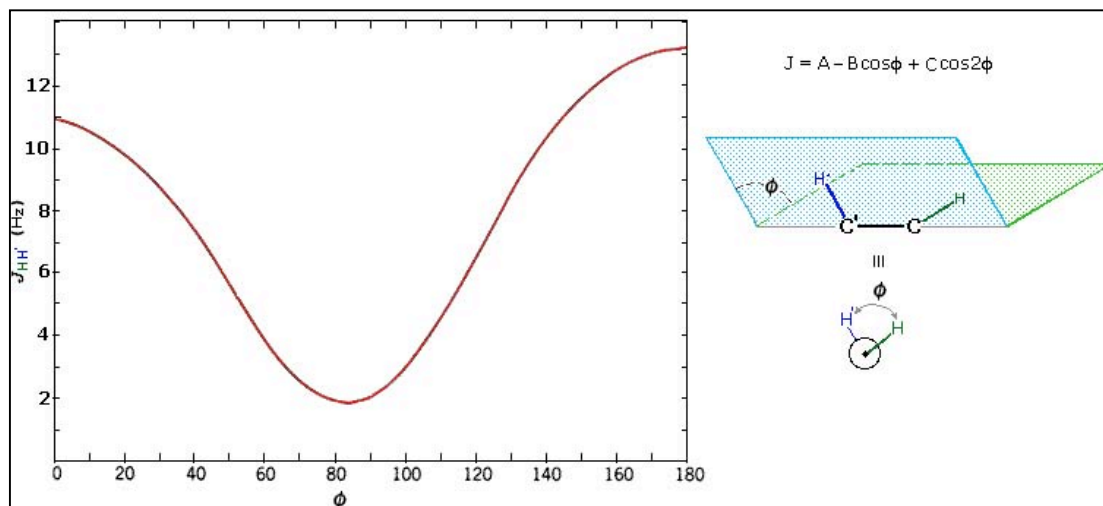
Figure 10: General form of the Karplus equation and correlation to the dihedral angles.

$${}^{3}J_{\text{C-H}} = 5.7 \cos^2(\phi) - 0.6 \cos(\phi) + 0.5 \qquad \textbf{(Eq. 1)}$$

$${}^{3}J_{\text{C-C}} = 3.49 \cos^2(\phi) + 0.16 \qquad \textbf{(Eq. 2)}$$

1.5.2.2 The bioactive conformation of sLe^x

The bioactive conformation of sLe^x (the conformation adopted upon binding to E-selectin) has been broadly investigated by NMR through the years. Cooke *et al.*, for example, concluded that the bound conformation was not identical as the one in solution, by comparing the transfer-nOe of sLe^x in water and in complex with E-selectin [162]. In contrast to these findings, Hensley *et al.* reported that the two conformations were indeed identical [163]. This discrepancy, further discussed by Scheffler *et al.* [164,165], arose from the presence or absence of the NOE between the H–3 of Gal and the H–3(ax) of Neu$N$Ac when examining the solution conformation of sLe^x (cf. above). This nOe is, however, absent in the bound conformation. In addition to that, in the bound state a nOe between the H–8 of Neu$N$Ac and the H–3 of Gal can be observed. By combining this two findings, it becomes clear that, in the bound state, the Neu$N$Ac-Gal linkage can only adopt the (−)-gauche conformation described by the dihedral angles {$\Phi = -70° \pm 10°$, $\Psi = 0° \pm 10°$}. Furthermore, with the exception of Poppe *et al.* [153], a general agreement is also found for the conformation adopted by the Le^x-core of sLe^x (*Table* 3). More recently the suppositions of Scheffler *et al.* [164,165] were confirmed, when the first crystal structure of E-selectin co-crystallized with sLe^x was published [166]. As it is shown in *Table* 3, the bioactive conformation determined by NMR and X-ray are rather similar. The bioactive conformation

proposed by Scheffler *et al.* [164,165] is presented in *Figure* 11. Particularities of this conformation include: the hydrophobic interaction of the methyl group of fucose with the β-face of the galactose moiety, the carboxyl function of NeuNAc perpendicular to the GlcNAc plane and the GlcNAc moiety playing the role of a three dimensional spacer. This particular arrangement permits to sLe^x to present its pharmacophores within a row along one side of the tetrasaccharide, therefore facilitating the binding to the relatively shallow hydrophilic cleft of E-selectin. The results of Scheffler *et al.* [164,165] have been the basis for the definition of the so-called bioactive window by Kolb and Ernst (see below).

Table 3: Summary of the data presents in the literature for the bioactive conformation of sLe^x.

Exp. Method	Φ_1 (°)	Ψ_1 (°)	Φ_2 (°)	Ψ_2 (°)	Φ_3 (°)	Ψ_3 (°)	Ref.
NMR	70	14	25	33	−58	−20	[153]
NMR	20	34	52	22	−70	8	[154]
NMR	38	26	39	12	−76	6	[164,165]
X-ray	41	22	34	16	−65	−12	[166]
NMR	29	41	45	19	−43	−12	[167]
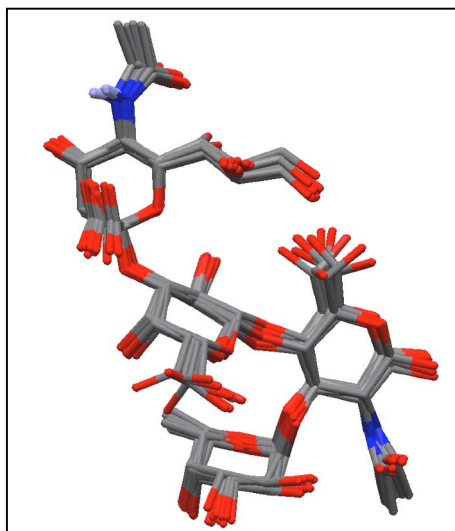
Figure 11: The bioactive conformation of sLe[x] as determined by Scheffler *et al.* [164,165].

1.5.2.3 Comparison of the bioactive conformation of sLe[x] in solution and in the bound state

By comparing the data of solution conformations presented in *Table* 2 and the one of the bioactive conformation in *Table* 3, it can be concluded how the Le[x]-core of sLe[x] undergoes no conformational change upon binding and seems to be rather rigid even in the free state. In contrast, different opinions reign for the conformation(s) adopted by the Neu*N*Ac-Gal linkage. It seems, however, probable that, in solution, a multi-conformational equilibrium exists and that, at least a part of the sLe[x] molecules in aqueous solution, have to adapt their conformation upon binding to E-selectin [150,162]. Experimental evidence for this statement will be presented in *Chapter* 4.2. The fact, that a multi-conformational equilibrium possibly exists, calls for designing conformationally-restrained compounds. Thus, an antagonist pre-organized in the bioactive conformation is thought to possess a higher affinity due to a more favorable entropic contribution than its torsionally less constricted counterpart.

The theoretical background for this concept is given by the fact shown in *Figure* 12, small differences in energy and/or different conformational distributions strongly affect the binding affinity of a compound towards its target protein.

$$\Delta G° = -RT \, lnK$$

Energy

$\Delta G° = 0.5$ kcal/mol	$\rightarrow$	57.6 % preference for minima A
$\Delta G° = 1.0$ kcal/mol	$\rightarrow$	81.5 % preference for minima A
$\Delta G° = 2.0$ kcal/mol	$\rightarrow$	96.6 % preference for minima A
$\Delta G° = 3.0$ kcal/mol	$\rightarrow$	99.4 % preference for minima A

Minima B $\Delta G°$ Minima A

Energy

$\Delta G° = 0.5$ kcal/mol	$\rightarrow$	affinity is reduced by a factor 2.3
$\Delta G° = 1.0$ kcal/mol	$\rightarrow$	affinity is reduced by a factor 5.4
$\Delta G° = 2.0$ kcal/mol	$\rightarrow$	affinity is reduced by a factor 29.3
$\Delta G° = 3.0$ kcal/mol	$\rightarrow$	affinity is reduced by a factor 158.8

Bioactive Conformation $\Delta G°$ Global energy minimum in solution

Figure 12: Theoretical background for the role played by conformations in drug design. Small energy differences between conformations translate in large differences in population. Moreover, if the less stable conformer is the bioactive one, affinity is reduced.

1.5.3 The computational model for the investigation of the conformational behavior of E-selectin antagonists

In 1997, Kolb and Ernst [161,168] validated a computational tool for the investigation of the conformational behavior of E-selectin antagonists and for predicting their affinity towards the target protein. The proposed protocol is based on the "Jumping between Wells" (JBW) simulation technique as implemented in *MacroModel* 5.0 [169]. In a first step, local minima conformations of the compound of interest are search by applying a systematic pseudo-Monte-Carlo (**s**ystematic, **u**nbounded **m**ultiple **m**inimum search, SUMM) method. The energetically most favorable 100 conformations thereby identified, are then used as an input for the following "Jumping between Wells" stochastic-dynamics simulation (JBW-SD). Thus, a Boltzmann-weighted ensemble of states is generated by jumping between different energy wells (the 100 minima retained

from the SUMM-search) and by then performing stochastic dynamics simulations within each well. All calculations were performed with an AMBER force field with parameters augmented for carbohydrates [161,168,170] and in conjunction with the GB/SA continuum-water model [171]. As a result, a distribution of the conformers of the mimic of interest in respect to the whole conformational space is obtained. More details on the simulation technique are given in *Chapter 3.2.3.1*.

To facilitate the data analysis of these simulations, Kolb and Ernst [161,168] also developed a 2D internal coordinate system (to be use instead of the three Φ/Ψ–plots for the three glycosidic bonds) that permits to define the spatial orientation of the relevant pharmacophoric groups of E-selectin antagonists (*Figure* 13). The two parameters chosen are the "torsion angles": Fuc(C4)-Fuc(C1)-Fuc(O1)-Acid(Cα) and Fuc(C1)-Fuc(O1)-Acid(Cα)-Acid(C=O). The first, describes the conformation of the Le^x core and is therefore referred to as *core conformation*. The second defines the orientation of the acid group relative to the core and is therefore referred to as *acid orientation*.
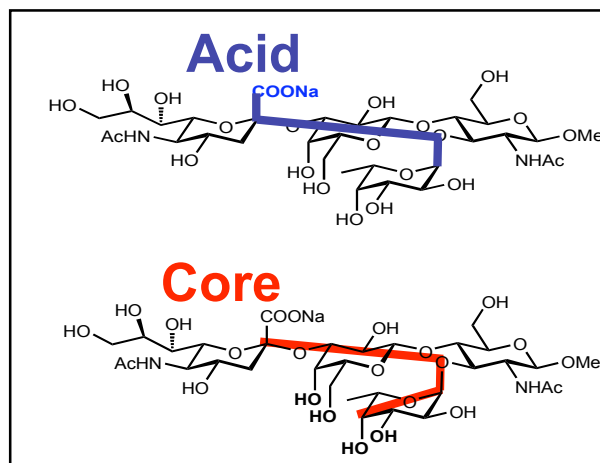
Figure 13: Definition of the acid-core "torsion" angles.

As can be seen in *Figure* 14, by plotting these two parameters against each other, a probabilistic distribution of the conformations of the compound of interest in the conformational space can be visualized.
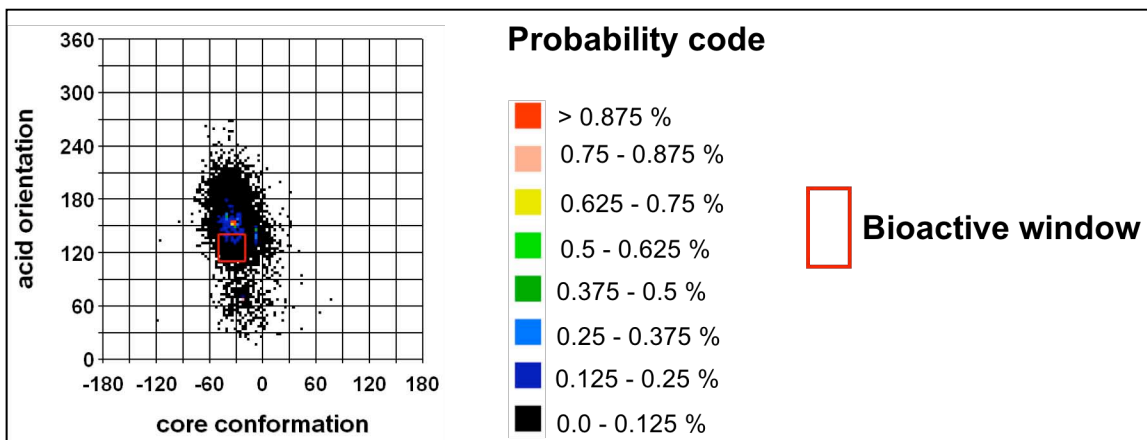
Figure 14: Probabilistic distribution of the conformation of a compound.

Moreover, this distribution can be compared to the one obtained with sLe^x (**3**) and the number of the conformations of compound of interest fitting the bioactive window, defined by Kolb and Ernst [161] (*Figure* 15) upon the results of Scheffler *et al.* [164,165], checked.
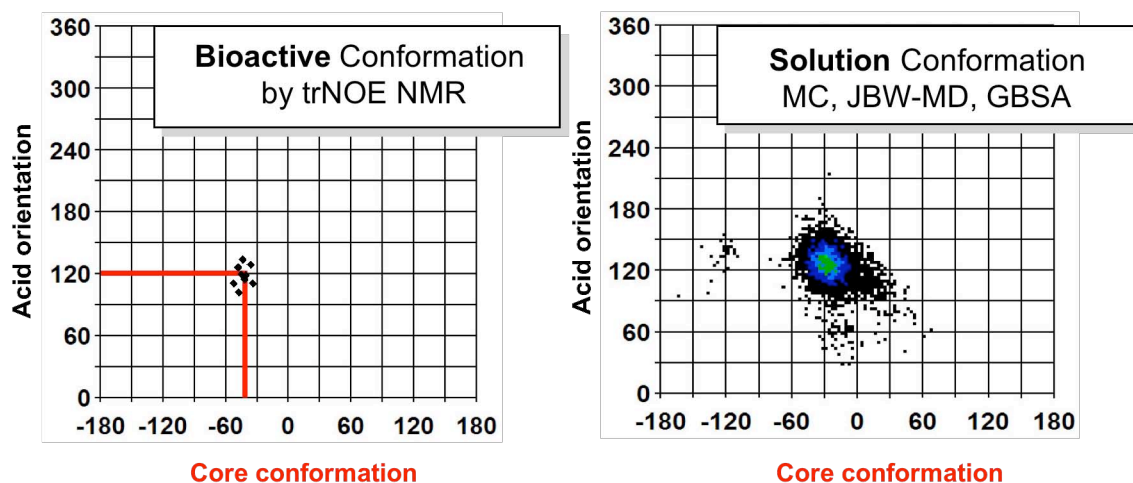
Figure 15: Definition of the bioactive window (left), based on the NMR of Scheffler *et al.* [164,165], and results obtained for the simulation o fsLe^x (**3**).

These two criteria are thought to give sufficient hints about the affinity of the compound of interest towards E-selectin. In fact, it was shown [161,168] that a compound featuring a strong pre-organization (therefore having a high percentage of conformations fitting the bioactive window in the JBW-simulation) and a more "concentrated" distribution of conformations compared to sLe^x,

usually has a stronger affinity to the target protein due to the reduced entropy costs to pay upon binding. The proof of concept for this approach based on the pre-organization of E-selectin antagonists, was given by Kolb and Ernst [161,168], who designed a series of compound, thereby showing that the model results were able to predict their biological activity (*Figure* 16). A powerful tool for the design of new sLe^x mimics was since then at disposition.
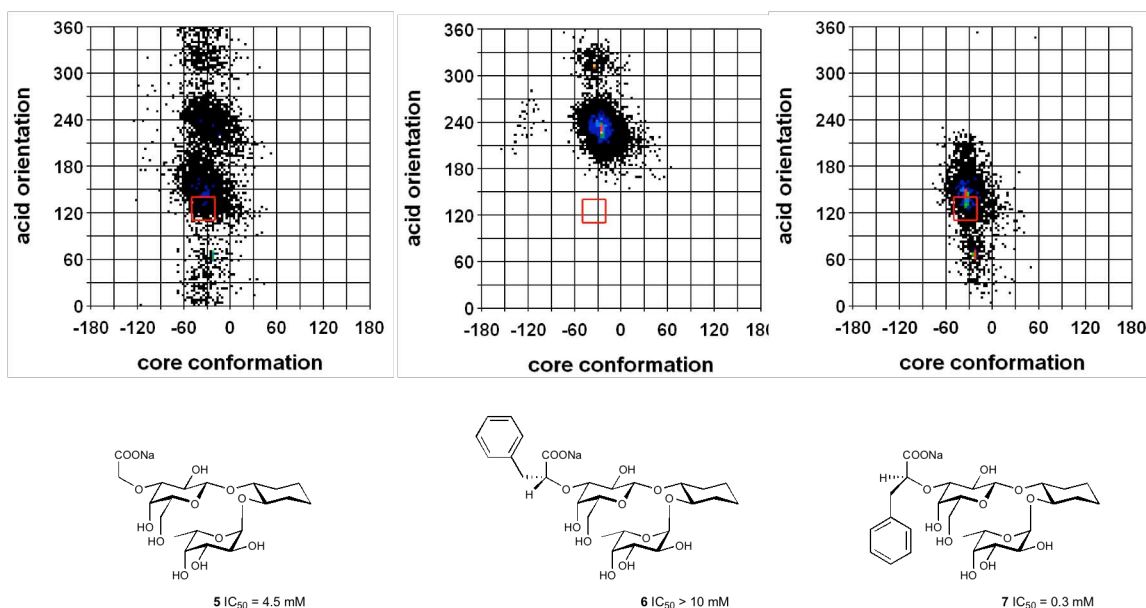
Figure 16: Chemical structure, simulated conformational variability in solution, and biological activity of the compounds **5–7**.

1.5.4 Hypothetical models for the binding mode of sLe^x to E-selectin

Different models, predicting the binding mode of sLe^x to E-selectin, have been proposed in the decades before the crystal structure determined [166] gave a detailed insight into the molecular recognition of sLe^x by the protein. It is interesting to note, that each model predicted some of the interactions between protein and ligand correctly, but that none of them has been able to reproduce the binding mode as observed in the crystal structure. In particular, all models failed to correctly determine the interaction of fucose with the calcium ion. This discrepancy is mainly due to the fact that all models based on the hypothesis that the binding of sLe^x to E-selectin would be analogous to the one of mannose in the mannose-binding protein A (MBP-A) but this supposition turned out to be incorrect [172].

The crystal structure of MBP-A complexed with a mannose-containing oligosaccharide was published by Drickamer *et al.* [84] in 1992. It has been the

first structure of a saccharide bound to a C-type lectin domain. The most interesting observation was the role played by the 3– and the 4–hydroxyl groups of mannose (both equatorial) in the complexation of the calcium ion. Due to the high degree of homology between MBP-A and the selectins, this crystal structure became unfortunately (see above), the basis for the development of all the models aming to describe the interactions between E-selectin and sLe^x.

The first model, based on mutagenesis studies, was presented in 1992 by Erbe *et al.* [92]. They hypothesized that sLe^x bound to E-selectin in a small shallow pocket, formed by the amino acids Arg97, Lys111, Lys113, Ser47 and Tyr48. The first true insight into the binding site of E-selectin was provided by Graves *et al.* [173], who solved the x-ray crystal structure of the E-selectin CRD/EGF domains. This structure presented some differences when compared to MBP-A. In particular, in E-selectin the calcium co-ordination sphere contains only four amino acid residues instead of the five observed in MBP-A. Further details on the binding mode of sLe^x to E-selectin delivered the crystal structure of sLe^x co-crystallized with a selectin-like mutant of MBP-A [172]. Hence, the binding of the 2– and 3–hydroxyl groups of the fucose to the calcium ion was confirmed but surprising, despite earlier finding (e.g. pharmacophore studies), no interactions between the carboxylate group of sialic acid and the protein were observed. Two theories were therefore put forward: either sLe^x binds differently to E-selectin than to the MBP-A mutant, or the importance of the sialic acid, demonstrated by the SAR-studies, is not directly related to protein binding.

Due to their historic importance the two models, presented by Kogan *et al.* [174] and Ernst *et al.* [175] respectively, will be briefly discussed. The model proposed by Kogan *et al.* (*Figure* 17), is based on the bioactive conformation of sLe^x proposed by Cooke [162], which was docked onto E-selectin after having pre-oriented the fucose moiety of sLe^x the way it had been found in the mutated MBP-A crystal structure [172]. In contrast, the model of Ernst *et al.* (*Figure* 17) is based on the docking of the bioactive conformation obtained from NMR investigations [164,165] onto the crystal structure published by Graves [173].
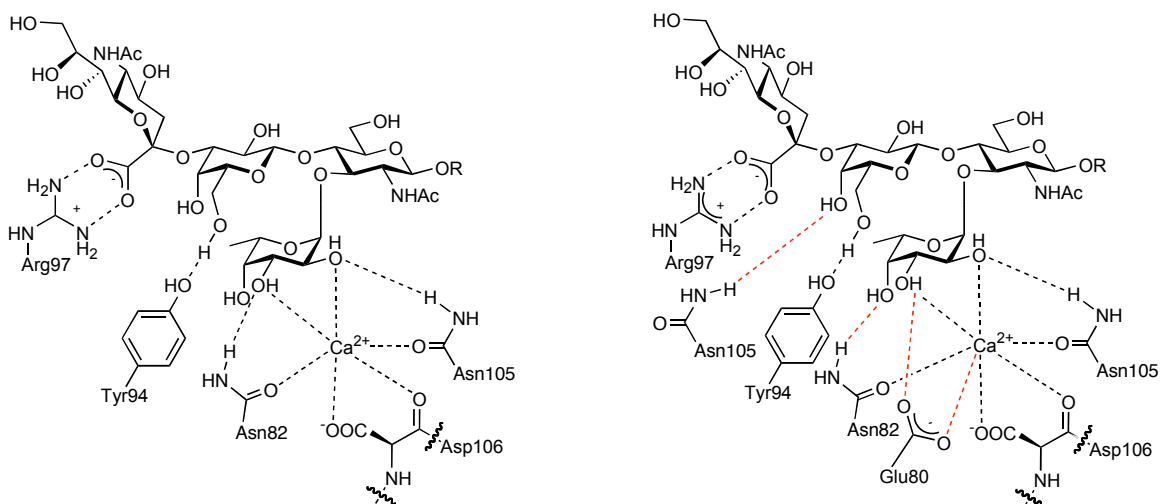
Figure 17: The model proposed by Kogan *et al.* [174] (left) compared with the one of Ernst *et al.* [175] (right). The differences are highlighted in red in the model by Ernst *et al.* [175] (right).

Both models suggest that the calcium ion is co-ordinated by the 2– and the 3– hydroxyl groups of fucose, that the 6–OH of the galactose moiety is interacting with Tyr94, and that the carboxylate group of the sialic acid residue is in close contact with Arg97. Slight differences lie in the amino acids coordinating the calcium ion, the contact of the 4–OH of Gal to Asn105 being not predicted by Kogan and the interaction of Asn82 with the hydroxyl groups of fucose. Recently, however, a directly insight into the molecular recognition of sLe^x by E-selectin was gained through the crystal structure of the E-selectin/sLe^x complex obtained by Somers *et al.* [166]. This structure will be discussed in *Chapter* 1.5.5.

A very different model, proposed first by Ernst *et al.* (personal communication) and further developed in this thesis (cf. *Chapter* 4.4.3), is based on the so-called reverse docking mode. This model may explain the finding of Defrees *et al.* [176], who observed that the introduction of hydrophobic substituents at the 2–position of the GlcNAc moiety leaded to higher affinity. In fact, as stated in [176] (”… when the synthesized compounds were incorporated into several proposed sLe^x/E-selectin binding models, specific interactions could not be identified …”) the reasons for this enhancement were unclear. To possibly identify these specific interactions, we screened the surface of E-selectin for hydrophobic regions, which lead to the proposition of the reversed binding mode illustrated in *Figure* 18.
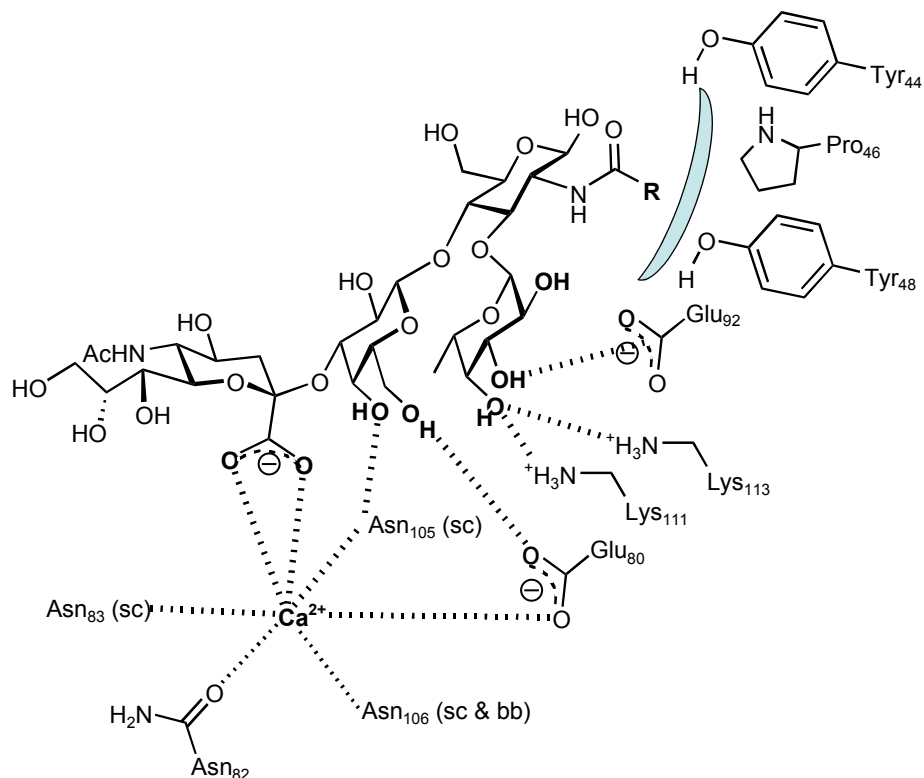
Figure 18: Schematic representation of the "reverse docking mode". The calcium ion is coordinated by the carboxylate of the sialic acid, whereas the hydrophobic substituents at the 2–position of the Glc*N*Ac moiety (**R**) interact with the hydrophobic region of the protein formed by Tyr44, Pro46 and Tyr48.

1.5.5 The crystal structure of the sLe^x/E-selectin complex: an insight into the "true" binding mode

Recently, Somers *et al.* [166] (*Figure* 19) significantly improved our understanding of selectin/carbohydrate binding by determining the crystal structures of complexes of sLe^x with the lectin/EGF domains of P- and E-selectin. Their work not only provided new insights, but also a few surprises. In particular, the majority of the contacts appear to be electrostatic in nature and the total buried surface of the complex is relatively small (only 550 Å^2) when compared to the size of the free ligand. Moreover, the selectin-bound calcium ion is complexed by the 3– and the 4–hydroxyls of Fuc and not, as suggested before by the 2–OH and the 3–OH. Further, the 4–OH of Fuc is also involved in hydrogen bonds with the residues Asn82 and Glu80, where as the 3–OH interacts, even with Asn105. The 2–OH of Fuc is, then again, indirectly bound to Asn83 and Glu107 through water-mediated hydrogen bonds. Looking at the Gal

moiety, contacts between the 4–OH and Tyr94, as well as between the 6–OH and Glu92 could be detected. The carboxylate group of the sialic acid binds the amino acids Arg97 and Tyr48.

The conformation adopted by sLe^x in the crystal structure is not analyzed in detail by Somers *et al.* [166], but, as it was already discussed above, is very similar to the bioactive conformation found by Scheffler *et al.* [164,165] by NMR and corresponds to minina C of the Ichigawa simulation [152] (cf. *Table* 2).
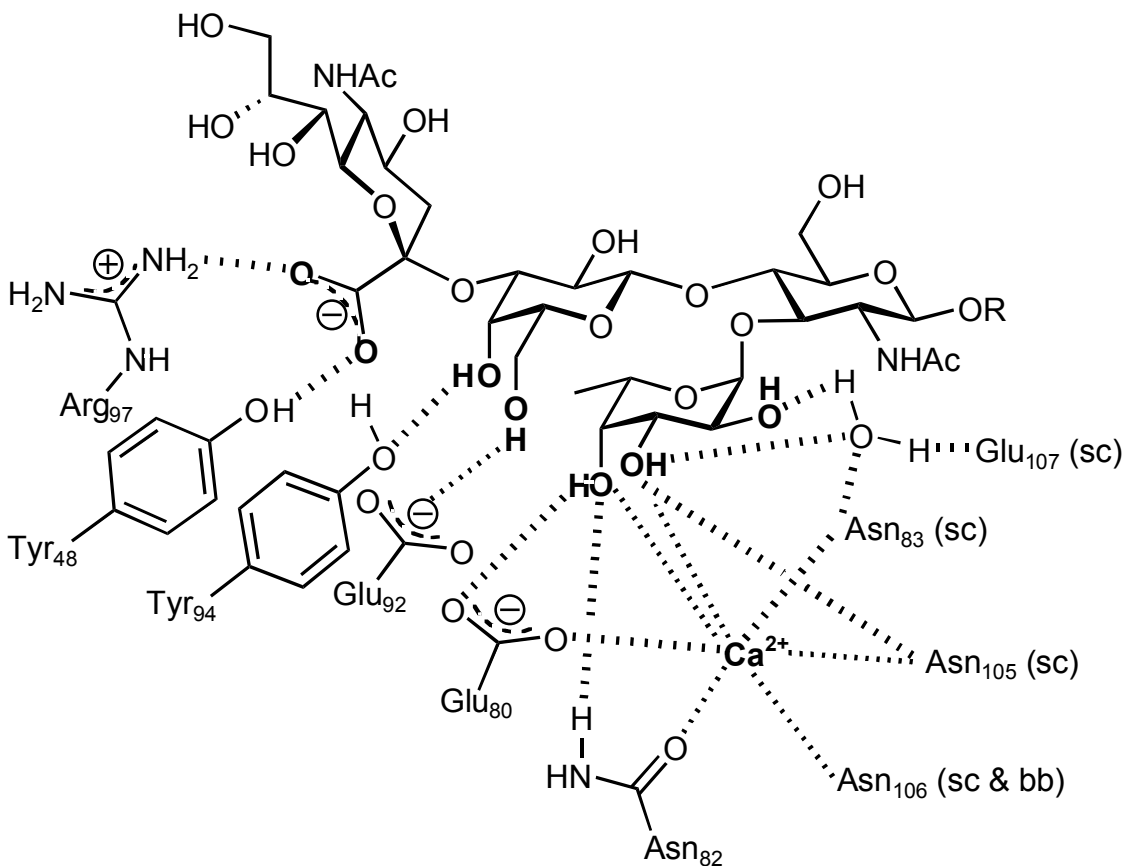
Figure 19: The binding mode of sLe^x as found in the crystal structure [166].

1.6 Towards the development of sLe^x mimics as selectin antagonists

After demonstrating that E-selectin plays a key-role in the inflammatory cascade, substantial interest arose on developing ligands that could act as anti-inflammatory agents. The terminal carbohydrate epitope sialyl Lewis^x served as a lead in the search of new E-selectin antagonists fulfilling two main criteria:

- modification of the lead structure towards molecules that overcome the pharmacokinetic (i.e. fast renal excretion, high polarity leading to low bioavailability) and pharmacodymamic disadvantages typical of carbohydrates (low affinity), and

- simplification of the complex structure of sLe^x.

Most studies aimed at substitute one or more sugars moieties of sLe^x with other structures, thereby retaining the key pharmacophores. A key aspect in this regard is the loss of the conformational pre-organization of pharmacophores, caused by the replacement of the relatively rigid Lewis^x-core structure by more flexible linkers. As it will discussed below, this kind of modifications often lead to less affine ligands, because of the high entropic cost associated with binding.

Another possibility explored to circumvent the problem of the low affinity of sLe^x towards E-selectin, was a polyvalent approach. This showed, at least in part, promising results [177–180], but these compounds are not suited for oral drug application. Several reviews cover this aspect [181–189].

This chapter will be dedicated to the most interesting sLe^x mimics developed to date. To facilitate the reading, the overview of the different contributions, is composed into sections corresponding to the number of sugar moieties replaced. Within each group (one sugar less, two sugars less,…) further strategies such as addressing secondary binding sites for additional interactions or pre-organization of the pharmacophores in the bioactive conformation by rigidification, will be discussed.

1.6.1 Trisaccharide mimics

1.6.1.1 Deletion of sialic acid

It is known from SAR-studies, that only the carboxylic acid of the neuramic-acid moiety contributes to E-selectin binding. The first logical step was therefore, to replace the costly and synthetic demanding sialic acid by a simple negatively charged group. For instance, sulfated Le^x trisaccharide (**8**) does exist in nature,

as a natural analog of sLe[x], and shows even a 20–fold superior binding affinity towards E-selectin [190][1]. Similar results were achieved by the substitution of the sialic acid moiety by a phosphate group (**9**) [191]. However, the most common simplification, is the alkylation of the 3–OH of Gal. The 3–carboxymethyl-substituted analog (**5**) shows similar affinity as sLe[x]. A more rigid NeuNAc mimic was synthesized by Duthaler *et al.* [192] (**10**). This compound, however, showed no affinity towards E-selectin. In fact, the acid moiety was fixed in a conformation different to the bioactive one. Other NeuNAc replacement will be presented later in combination with GlcNAc-substitutions.
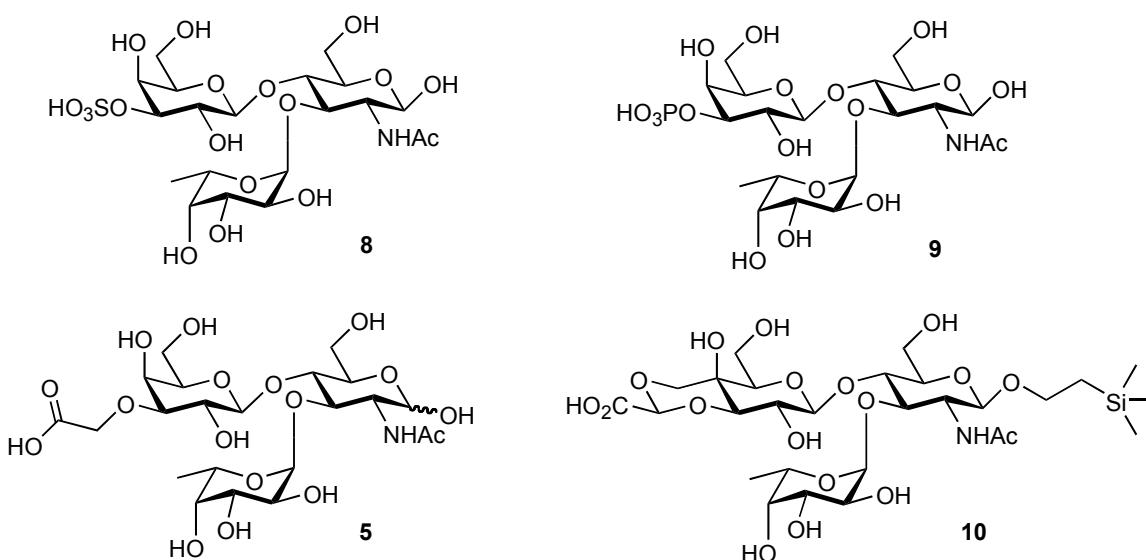
Figure 20: Mimics containing a negatively charged group replacing NeuNAc.

1.6.1.2 Substitution of the GlcNAc-moiety

As it was confirmed by the crystal structure of the sLe[x]/E-selectin complex [166], the GlcNAc-moiety of sLe[x] seems not to be directly involved in any interaction with the protein. However, this moiety plays an important role as a three dimensional spacer correctly orienting the pharmacophores. Therefore, several compounds have been synthesized, where the GlcNAc-moiety has been replaced by synthetically less demanding groups and/or by groups with a better pharmacokinetic and pharmacodynamic profile. Hannesian [193] for example replaced the GlcNAc-moiety by an indolizidinone unit. This structure (**11**) showed no affinity towards E-selectin anymore but bound P-selectin stronger than sLe[x] [193]. Substitution of GlcNAc with quinic acid leads to compound **12**, which was

---

[1] However, these finding could not be reproduced by numerous other groups active in the selectin field. (Beat Ernst, personal communication)

equally active as sLe^x [193]. However, the most successful and, from a pharmacokinetic point of view, interesting substitution was obtained by Töpfer [194]. He replaced GlcNAc by a (*1R,2R*)-cyclohexandiol ring. His compound (**13**) was three times more active than sLe^x.
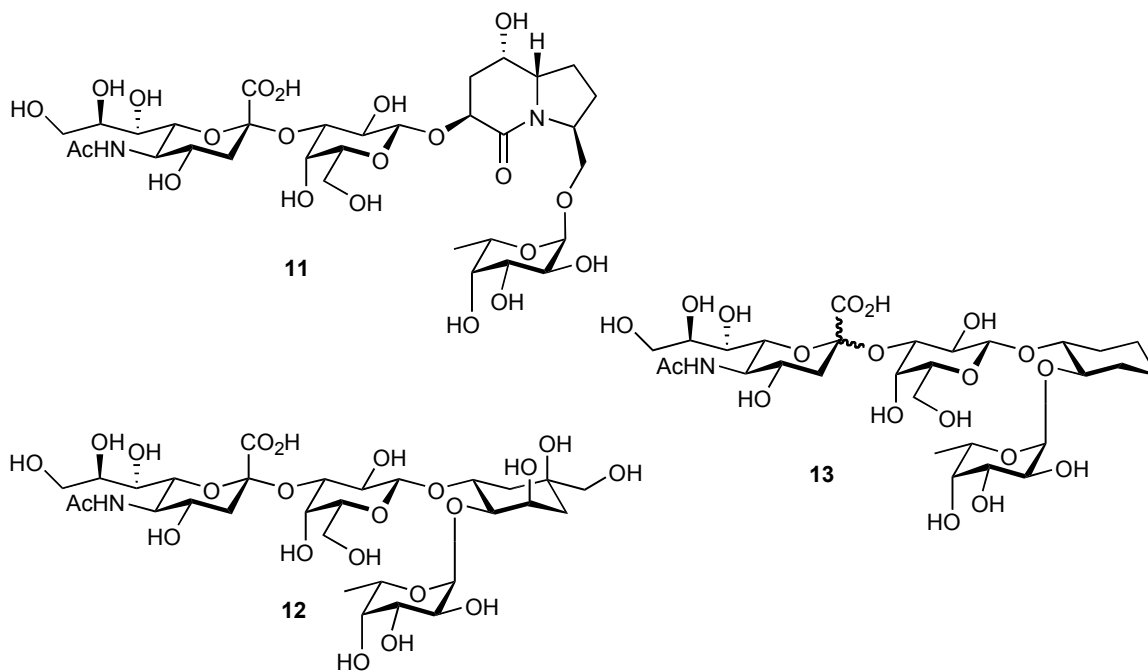
Figure 21: Mimics containing different GlcNAc replacement.

1.6.2 Two-sugar mimics

1.6.2.1 Replacement of NeuNAc and GlcNAc

The next step in the simplification process of sLe^x was to combine the effects of the NeuNAc- and GlcNAc-replacements. The substitution of GlcNAc solely, hardly improved binding affinity but lead to mimics easily to be synthesized and with improved pharmacokinetic-properties. A variety of GlcNAc-replacements have been tested in combination with glycolic acid or alkyl- and aryl lactic acid residues mimicking the sialic acid part (*Figure* 22 and 23).
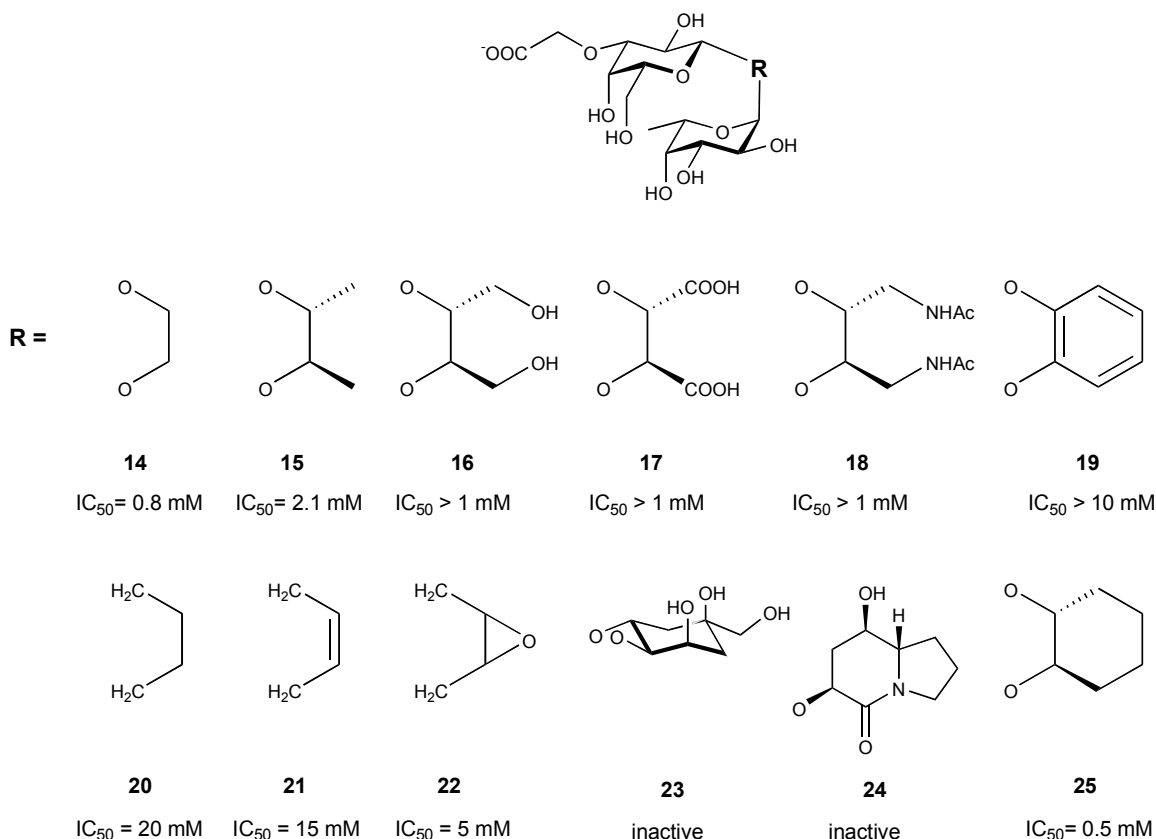
Figure 22: sLe^x mimics bearing a carboxymethyl group as sialic acid replacement and a variety of more or less flexible linker groups substituting the GlcNAc unit.

Wong suggested [182], that a substitution of the GlcNAc-moiety by a (*1R,2R*)– cyclohexandiol ring, seems to be energetically neutral when, in parallel, NeuNAc is exchanged against a carboxymethyl group (**25**). Surprisingly, compound **14**, which contains a much more flexible linker leading to a reduced stability of the Le^x core structure, was almost as active as **25**. In general however, the introduction of flexible 1,2 diols as GlcNAc-mimics lead to a reduced affinity of the ligands towards E-selectin (**15–18**) [195]. Further modifications included the introduction of a butane- (**20**), a cis-olefin- (**21**) or an epoxide-linker (**22**) between Fuc and Gal, but none of these GlcNAc-substitution led to higher affinities. A probable explanation for the lost of affinity with this class of compounds lies in their lack of a rigid linker between Fuc and Gal, stabilizing the Le^x-core structure. In addition, the mimics containing quinic acid (**23**) or indolizidinone (**24**) as a GlcNAc-replacement (even in combination with the positive carboxymethyl substitution of NeuNAc) didn't show any improvement in affinity compared to sLe^x.

Of particular interest for the work performed in this thesis project and at the Institute of Molecular Pharmacy in general, are the mimics developed by Ernst *et al.* [161,168,196]. By using a molecular-modeling tool to be described later in detail, they identified (*2S*)–cyclohexyl- and (*2S*)–phenyl-lactic acid as a plausible simplification of the Neu*N*Ac-moiety and further established modifications of the (*1R,2R*)–cyclohexandiol ring initially proposed by Wong [182] as a valuable Glc*N*Ac-replacement. Regarding the sialic acid substitution, they could show that the introduction of sterically demanding groups next to the acid function, forces the carboxylate to adopt the bioactive conformation observed for sLe^x. In fact, compound **26** that will often be used for reference purposes in this thesis, turned out to be ten to twelve times more active than sLe^x [161,168]. Obviously, the *R*-stereoisomer was found to be inactive (**27**) [161,168]. Even more active compounds (**28-32**) were obtained by combining the Neu*N*Ac-substitution through (*2S*)–cyclohexyl-derivative with the introduction of aliphatic or aromatic substituents at the position two of the Glc*N*Ac mimic [196]. The role played by these substituents will be part of the investigations performed in this thesis. A further improvement in affinity was obtained by serendipity by Thoma *et al.* [197]. Due to an incomplete deprotection of the galactose, compound **36** was obtained and turned out to be three times more active than the corresponding mimic with a free 2–OH group. A rational explanation for this effect will be hypothesized below (cf. *Chapter* 4.7.2.2).

Various modifications at the 6-position of Gal of compound **26**, as its rigidification leading to compound **35** didn't lead to any improvement in binding affinity [149].
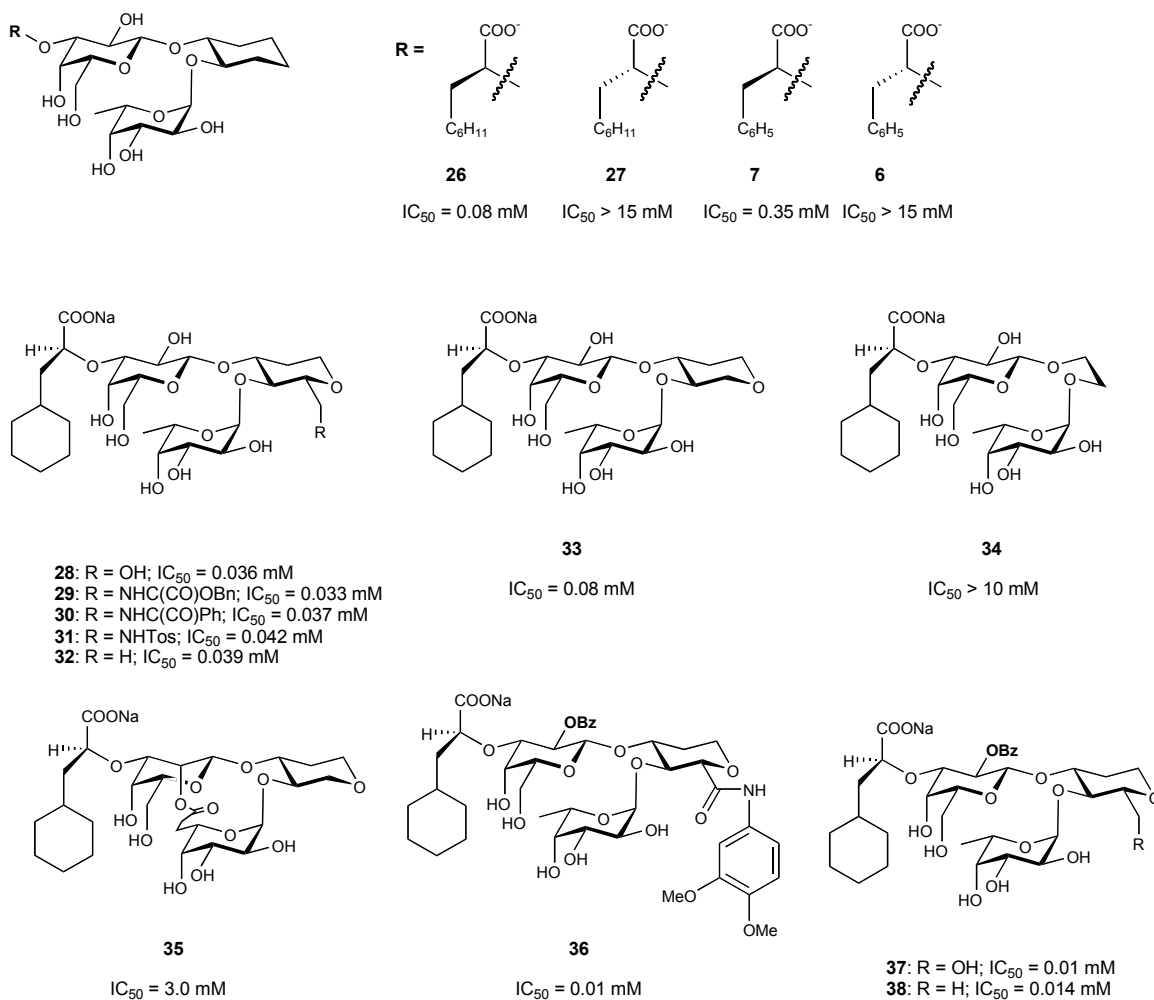
Figure 23: Mimics developed at Novartis AG, Basel.

1.6.2.2 Replacement of the *N*-acetyl-lactosamine disaccharide

Another approach explored in the search for potent E-selectin antagonists, is the replacement of the central *N*-acetyl-lactosamine disaccharide unit (*Figure* 24). The introduction of unfuctionalized linkers (**42–44**) [198] instead of the disaccharide unit, however delivered only poor results. Reasonable explanations for this lack of activity are the high entropic costs resulting from the higher flexibility and the lack of functionalities imitating the essential pharmacophores of Gal (4–OH and 6–OH). To circumvent this second problem, Töpfer *et al.*[194] proposed a propanediol-cyclohexan linker (**39–41**), but no breakthrough has yet been achieved. Also the use of more complex and rigid linkers like an inflexible spiroketal scaffold (**47**) [199] or a benzenedimethanol (**45**, **46**) [200] didn't fulfill the expectations and led only to poor-affinity ligands.
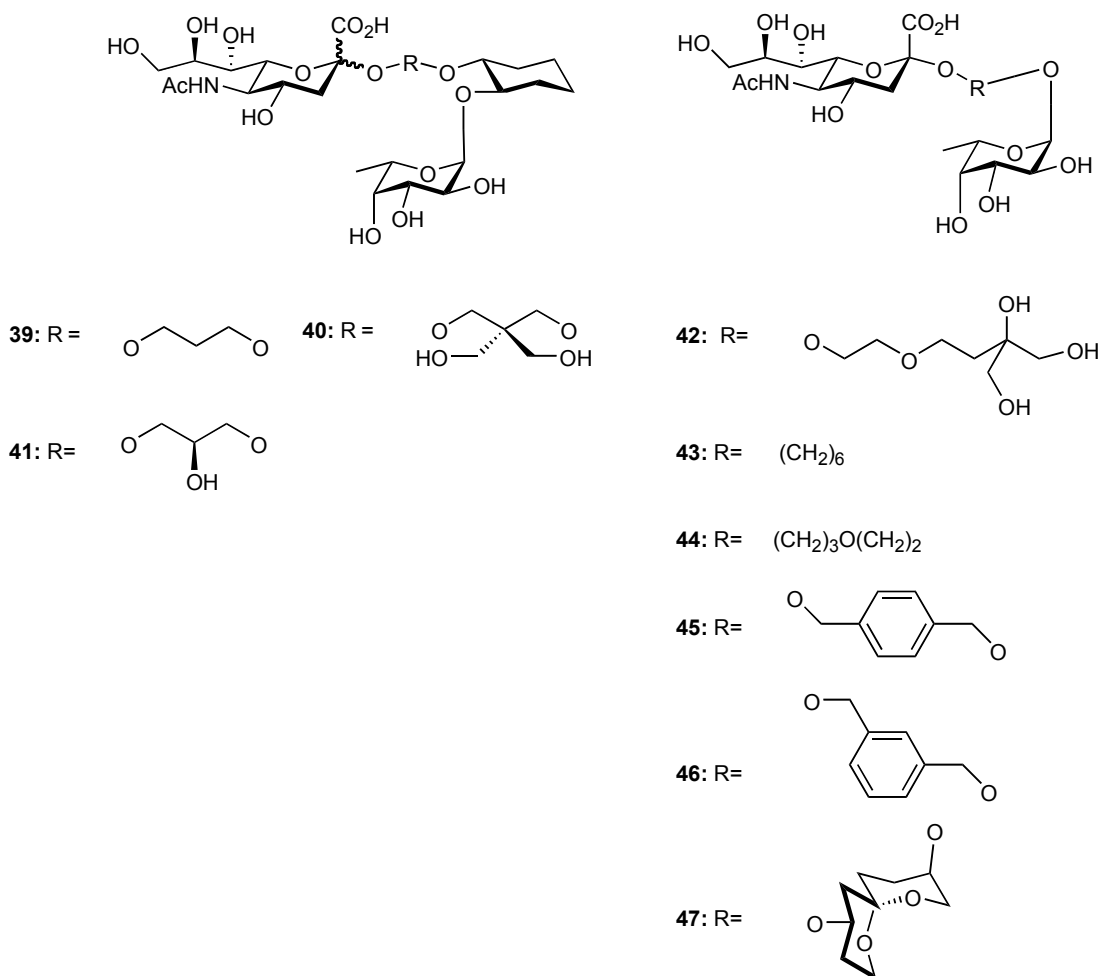
Figure 24: Mimics in which the N-acetyl-lactosamine disaccharide has been substituted.

1.6.3 One sugar mimics

The fucose moiety of sLe^x, which coordinates the calcium ion in the selectins and contains three of the six pharmacophoric groups, is probably the most difficult residue to substitute. Therefore, the largest and more "drug-like" group of selectin antagonists is based on structure containing only this sugar unit, in some cases replaced by mannose or L-galactose. Additional functional groups were then attached to this moiety to mimic the omitted pharmacophoric groups of the Gal- and Neu*N*Ac-residues. A series of mimics integrating the established linker (*1R,2R*)–cyclohexandiol ring as a Glc*N*Ac replacement had some success (*Figure* 25). For instance, Töpfer *et al.* [201] used this functionality to prepare a series of mimics with malonic acid derivatives (**53–62**). Bänteli *et al.* [202] and Liu *et al.* [203] prepared fucose-based mimics featuring different disubstituted aryl groups to mimic the Neu*N*Ac-Gal moiety (**48–51**) but all the substances were found to be inactive. The authors suggested that the use of an aromatic spacer

instead of Gal probably does not allow the steric pre-organization needed for binding. Furthermore the aromatic substitution leads to a completely different geometry (the aromatic ring is planar vs. the chair conformation of Gal), making it impossible to appropriately position the pharmacophoric groups.
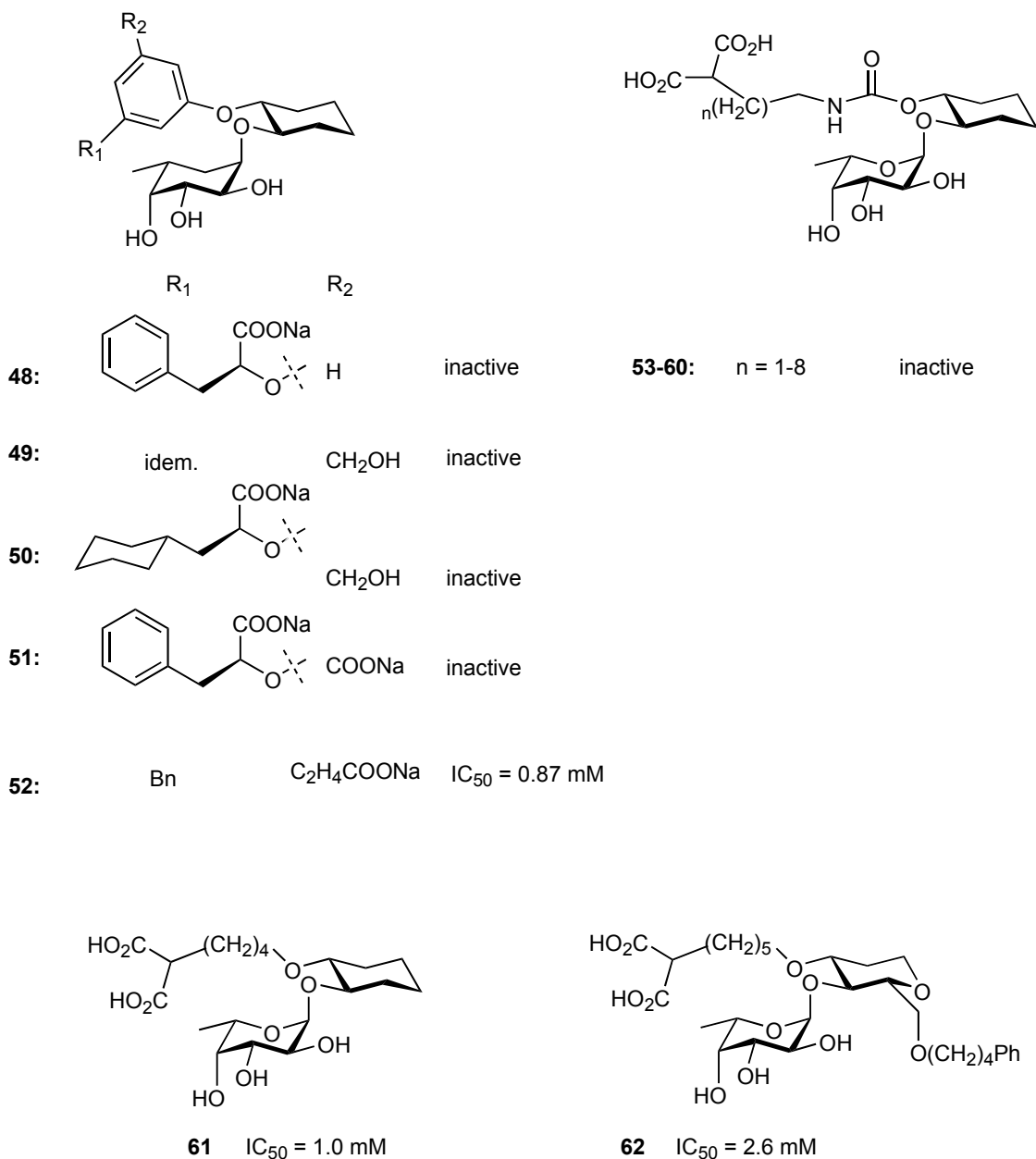
Figure 25: Fucose-based sLe^x-mimics containing a cyclohexanediol spacer combined with a variety of linkers to attach the anionic endgroup.

Kogan *et al.* [204,205] developed a library of different glycoaromatics, based on a derivatized biphenyl residue attached to the anomeric position of mannose (**i**–**iii** *Figure* 26). With the exception of compounds **63**-**65**, the ligands displayed no affinity towards E-selectin. However, nine molecules showed a two to twenty fold better activity towards P-selectin than sLe^x. In addition, they studied dimeric glycoaromatics to mimic extended sialyl di-Lewis^x structures isolated form human neutrophils [121]. Dimer **65** was six times more active than sLe^x against E-selectin in a cell-based assay. **65** is currently in phase IIa clinical trials conducted by Revotar AG [206] for the treatment of asthma and psoriasis.

The group of Wong [207–212] focused its interest in generating a large library of fucose-, mannose-, or galactose-based glycopeptides. Two design elements were chosen as variables: "turns" mimicking the Glc*N*Ac unit, and "hydroxyls" mimicking the galactose pharmacophores. In general, those compounds showed a distinct improvement in affinity for P-selectin and usually only moderate for E-selectin. It is, however, difficult to derive a SAR-profile for these glycomimics due to the substantial differences among the structures and in the range of activity (from no activity to μM). Nevertheless, this series contains the most active inhibitors known to date. Then again, none of them has ever entered clinical trials. A detailed discussion of these compounds can be found in a recent review [182].
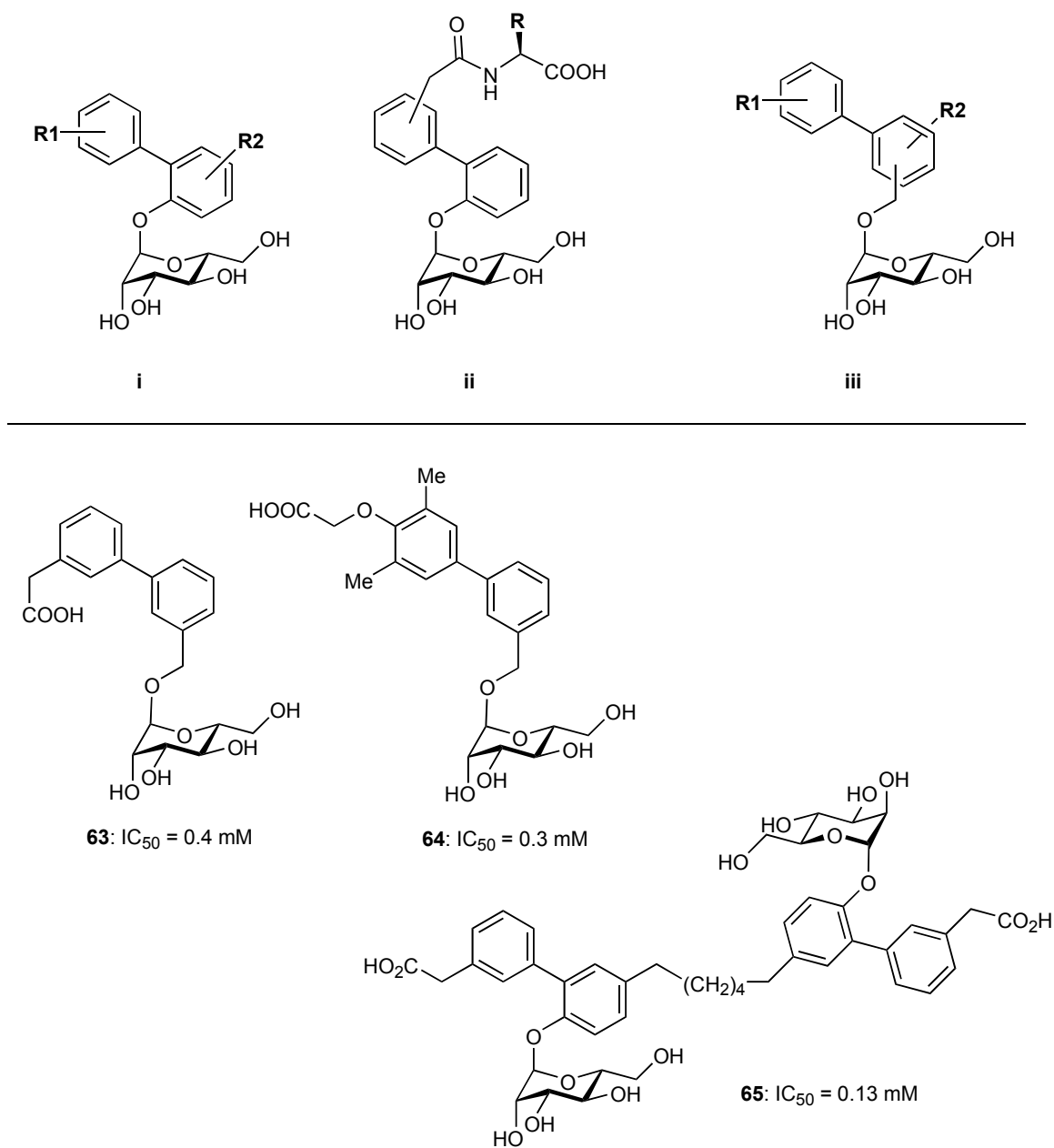
Figure 26: Biphenyl-based inhibitors investigated by Kogan *et al.* [204,205].

1.6.4 Groups addressing secondary binding sites

The compounds summarized so far, generally tried to improve binding through the simplification of the lead structure sLe^x. Hence, this approach is based on maintaining the pre-organized pharmacophore and on removing of the non-pharmacophoric functional groups, so that a benefit on the entropic level can be achieved. This approach led, in some cases, to more potent ligands, but a real breakthrough was not achieved. To further improve the activity of selectin antagonists, various attempts have been made to access additional (preferably hydrophobic) interactions with the receptor without loosing the key contacts described above (cf. *Chapter* 1.5.1). DeFrees *et al.* [176] showed that higher affinities are obtained when the acetyl group of the Glc*N*Ac moiety of sLe^x is replaced by benzoate (*Figure* 27, **66**) or naphthoate (*Figure* 26, **67**).
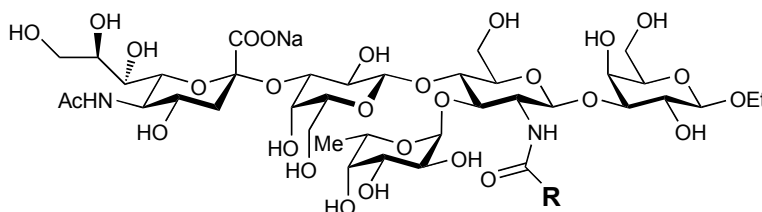
66 (R= benzoyl)
67 (R=naphthoyl)

Figure 27: The compounds identified by DeFrees *et al.* [176].

Ernst *et al.* [175] also investigated the effect of aliphatic, aromatic and heteroaromatic acyl substituents at the Glc*N*Ac-nitrogen. Their findings confirmed the results presented by DeFrees and provided even more potent ligands (*Figure 28)*. It is noteworthy that structure **68j**, with an activity 60 times (IC$_{50}$= 0.013 mM) higher than sLe^x, is one of the most potent inhibitors known to date. However, the structural basis for this improvement in activity is not fully understood, despite the availability of the crystal structure of the sLe^x/E-selectin complex [166]. This will therefore be analyzed in this thesis.

Moreover, the introduction of similar hydrophobic substituents to increase the affinity towards E-selectin, such as the ones proposed by Ernst *et al.* [175] for the 2–position of the Glc*N*Ac moiety, was hypothesized, as a part of this thesis, for the 2– and 6–positions of Gal and for the group neighboring the carboxylate functionality binding Arg97 and will be discussed in *Chapter* 4.7.2.
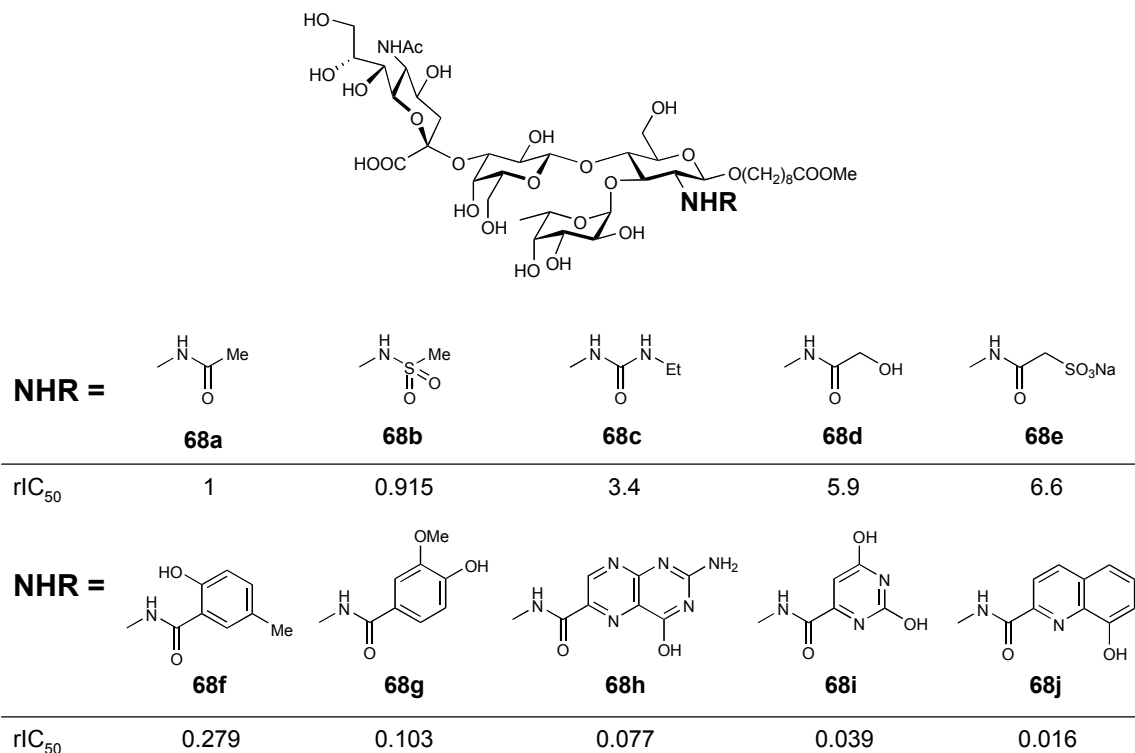
Figure 28: Modification of the 2–position of the GlcNAc unity by Ernst *et al.* [175].

1.6.5 Non-sugar based mimics

An alternative approach in the search for a high affinity ligand for E-selectin was undertook by Kondo *et al.* [213] and Brandley *et al.* [214]. They build up a pharmacophore model based on the information coming from SAR studies. They then screened databases looking for substances fitting to their pharmocophore hypohesis. Finally, they identified ligand **69** and **70** (*Figure* 29) as possible E-selectin antagonist. **69** showed an activity of 150 μM, **70** of 500 μM.

A further group of E-selectin inhibitors is formed by the peptidic antagonists. However, this class of antagonists will not be discussed in this thesis. For additional information please consult reference [215].
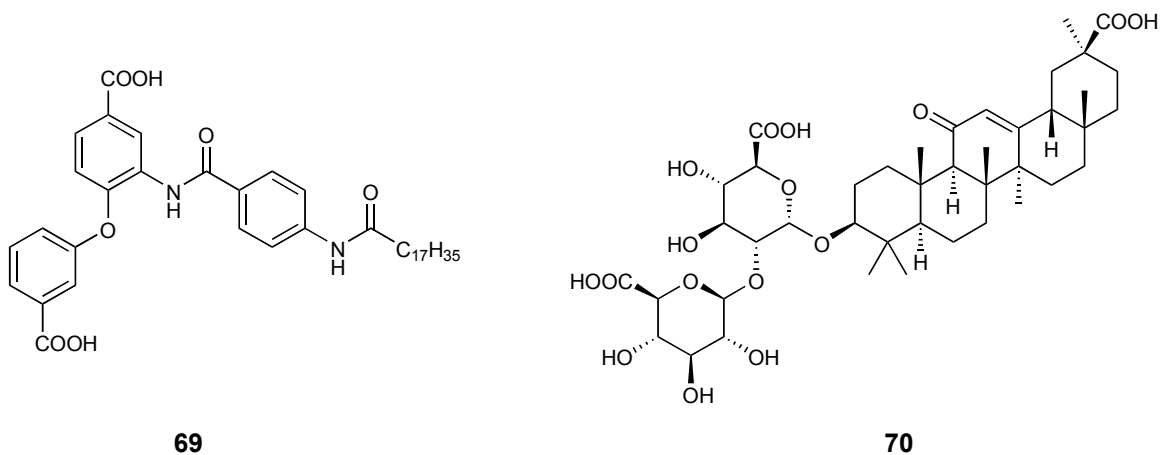
Figure 29: The two ligands identified by Kondo *et al.* [213] and Bradley *et al.* [214] through pharmacophore modeling.

1.7 Introduction to the modeling techniques

This thesis mainly deals with theoretical studies of E-selectin antagonists, therefore the following section will be devoted to a brief historic overview of the methods used to address the different topics related to the design and investigation of E-selectin antagonists. More details over the approaches chosen are given in the *Chapter* 3 and 4.

1.7.1 Conformational analysis

Molecules are not rigid. The thermal energy at 25°C is large enough to let all atoms in a molecule move permanently. Therefore, each compound containing one or more single bonds exists in many different so-called conformers or rotamers. The quantitative and qualitative composition of this mixture is permanently changing. Of course, the low-energy conformers contribute most to the composition of this mixture. A transformation from one conformation to another is usually related to changes in torsion angles around sterically unrestricted bonds. The changes in molecular conformations can be understood as movements on a multi-dimensional surface describing the potential energy and the geometry of a molecule. Each point on this surface represents the potential energy of a single state. Stable conformations correspond to the local minima on this energy surface. The relative population of a conformation depends on its statistical weight, which is both influenced by the potential energy and the entropy [216]. As a consequence, the global minimum on the potential energy surface is not always the most frequently populated. Conformational analyses are of particular interest, because it is assumed that the biological

activity of a drug molecule depends on one single unique conformation (the bioactive conformation) hidden among all the low-energy conformations [217]. The search for the bioactive conformation has therefore become one of the most challenging tasks in medicinal chemistry. In fact, as it can be observed in *Figure* 12 (cf. *Chapter* 1.5.2.3), small differences in $\Delta G°$ have a substantial influence on the conformation population, and in particular case, where the bioactive conformation doesn't coincide with the global minimum, on the binding affinity.

In the past decades, different theoretical techniques have been developed aimed at identifying low-energy conformers on the potential-energy surface [218–225]. In this work, only molecular-mechanics concepts will be discussed while QM-methods remain excluded.

1.7.1.1 Conformational analysis using systematic search procedures

The systematic search [219,220,226] is probably the most intuitive method. It is performed by systematically varying each of the torsion angles of a molecule in order to generate all possible conformers. Typically a step size of 15° per torsion is used. That means, that to sample a single torsion, 24 conformations have to be generated. Consequently, the number of conformations to be sample depends on the number of torsion (n) and on the step size (s) as given in *Equation* 3.

Number of conformations = $(360/s)^n$ (**Eq. 3**)

The disadvantage of this protocol is the huge amount of conformers that have to be sampled. If for example a molecule with six rotatable bonds would be investigated, by using a step size of 15°, the number of conformers generated, would already be of 191'102'976. In the past but even today, the calculation of the potential energy for each of this structure would be too time-consuming, and therefore ways to reduce the amount of conformations to be evaluated, were introduced. The first approach to reduce the data is to perform a van der Waals screening or "bump check". If two atoms are closer than the sum of their van der Waals radii the conformer would be rejected prior to the calculation of its energy. For the conformers that pass this first filter, the potential energy is computed using a molecular force field method. As a second filter, an energy window can be defined. The underlying idea for applying an energy window is based on the fact that conformations having much higher energy compared to the global minimum will be only neglectable populated (see again *Figure* 12). The value for this energy window depends on the size of the molecule studied and on the

applied force field. Typical values vary from five to fifteen kcal/mol above the global-energy minimum [224–228][2].

The structures that pass both filters should represent the complete ensemble of energetically accessible conformations of a particular molecule. However, in many cases the number of structures obtained, is still too large to be reasonable handled. Therefore, further techniques like clustering or factor analysis have been developed. For a discussion of these methods a detailed review is available [229].

1.7.1.2 Conformational analysis using Monte-Carlo methods

A completely different approach for screening the conformational space is based on Monte-Carlo search techniques. Random search techniques are of statistical nature [230]. At each step of the simulation, the actual conformation is randomly transformed in a new one by changing one or more values of the torsion angles (internal space refinement) [224] or by modifying the cartesian coordinates (cartesian space refinement) [221,222]. After each step the molecule is minimized and compared with the already sampled conformations. Only if it is unique, it will be retained. After this check, the randomization process is resumed. Potentially, the random methods allow to cover all the regions of the conformational space, but this is only true if the simulation runs for a sufficient time span. Hence, the probability to sample a unique conformation decreases dramatically depending on the growing number of conformers already discovered. Therefore, unless the computational time invested is indefinite, a residual probability of not having sampled all the conformational space remains. Hence, it is important to check the completeness of the analysis by performing different parallel runs starting the search protocol with different input conformations. The main advantage of a random search method is that, in principle, molecules of any size and also including ring systems can be successfully treated. In reality, however, highly flexible molecules like the carbohydrates investigated in this thesis, often don't give exhaustive sampled results, because the volume of the respective conformational space is too large. A slightly modified random method (SUMM searches) was therefore used to generate some of the results presented in this thesis. The difference to a standard Monte-Carlo approach relies on the fact that this method keeps track of the regions of the conformational space already visited, and thereby addresses the search towards new regions. A detailed description of this method can be found in reference [231].

---

[2] If this filter is applied, the actual solutions of the search have to be intermediately saved and resorted by energy after each simulation cycle.

1.7.1.3 Conformational analysis using Molecular-Dynamics Simulations

The aim of a molecular dynamics simulation approach is to reproduce the time-dependent motional behavior of a molecule. Basically, it is assumed that the atoms of a molecule move according to the Newton's laws, thereby interacting with each other through molecular forces. The simulation starts by applying a random velocity (Boltzmann distribution at a defined temperature) to each atom of the molecule of interest. At regular time intervals (e.g. 1.5 fs) then, the second law of Netwon is solved by integration, thereby yielding a new position and velocity of each atom of the system, the result being a new conformation. This cycle can then be repeated for a predefined number of steps. The collection of the energetically accessible conformations produced by MD is referred to as "ensemble". One of the most interesting features of this method is its ability to overcome energy barriers. Hence, by performing an MD, it should be possible to find also the minima other than the nearest on the potential energy surface (e.g. the global minimum). However, if the energy barriers are high or the number of degrees of freedom in the molecule very large, then some of the conformers can not be produced during the simulation time. Moreover, in the case of molecules with many rotatable bonds, due to huge conformational space that has to be sampled, the completeness of the search cannot be reached in a reasonable amount of time. To overcome the barrier problem, an often-used tactic is to perform the simulations at elevated temperatures, thereby introducing more energy to the system and thus facilitating the overcoming of energy barriers. A well-established protocol to perform conformational analysis by MD is to integrate this philosophy referred to as simulated annealing [232]. Here, the molecule is first "warmed up" to an elevated temperature (e.g. 600 K), and then, after equilibrating the system for a given period of time (e.g. 150 ps) at the elevated temperature, cooled down to the defined sampling temperature (e.g. 300 K) and thereby trapped in the nearest local minimum conformation. This conformation is then used as starting point for the "next high temperature cycle". In order to obtain a set of low-energy conformations, the heating-, equilibrating-, cooling-, and sampling-cycle should be repeated several times. This approach was mainly used for this work and will be presented in detail later (cf. *Chapter* 3.2.3.2).

1.7.1.4 Conformational analysis of carbohydrates

Due to the intrinsic characteristics of carbohydrates, their conformational analysis differs in several ways from that of the other organic molecules. The potential information content of carbohydrates is several orders of magnitude higher than any other biological macromolecule (e.g. peptides). The diversity of carbohydrate structures relies on the broad number of different monosaccharides available (>100) and on the different ways they are branched. For example, the number of all possible linear and branched isomers of a hexasaccharide containing only one

type of monosaccharide moiety exceeds 10^{12}, whereas only one hexapeptide is possible using always the same amino-acid as a repeated unit [233]. In addition to that, carbohydrates are highly flexible and will therefore often populate numerous conformations co-existing in solution even at room temperature. Furthermore, this floppiness represents a major problem in crystallizing carbohydrates, thereby rending the access to experimental data difficult. Finally, even though most recent development in NMR-techniques permits a better understanding of their structures, carbohydrates generally show only few inter-residual nOes in contrast to proteins. As a consequence, due to the lack of measurable data, it is often difficult to uniquely determine the conformation of an oligosaccharide through experimental methods [234]. However, also for theoretical methods this task remains difficult. Next to the huge number of conformations that has to be sampled, carbohydrates are especially though to model because of their highly polar functionality, the role of solvation, and the particular electronic effects present only in carbohydrates such as the anomeric-, the *exo*-anomeric- and the gauche-effect. The particularity of carbohydrates therefore induced the development of various force fields that may be suited for studying carbohydrates. Two extensive reviews on this topic have been published [235,236].

1.7.2 Docking and scoring procedures

1.7.2.1 An insight into the docking procedures

Docking procedures aim to identify the correct orientation/conformation of a small molecule in the binding pocket of a protein and to predict its affinity towards the target protein. Molecular docking has contributed important proceedings to drug discovery for many years and has become an essential tool in structure-based drug design. It is widely applied and meets very heterogeneous demands. Of course, the major task remains the identification of new active compounds for a particular target protein. Docking proved to be a reliable and fast filter in high-throughput virtual screening [237,238], thereby providing a pool of ideas for novel lead structure, and did produce several success stories [239–242]. In addition, docking is often used to predict the binding mode of new compound classes for which no crystal structure has been determined yet [243–247].

The setup for a ligand docking approach requires the following components: a target protein structure with or without a bound ligand, the molecules of interest and a computational framework that allows the implementation of the desired docking and scoring procedures. To guarantee a certain accuracy, the three-dimensional structure of the protein-ligand complex has to be detailed at atomic resolution. Most of the docking algorithms assume the protein to be rigid, according to the high computational cost that the demand of flexibility implicates (in some cases, like in *Yeti* [248,352], which also allows for dynamic solvation of

the binding pocket, flexibility of the amino-acids sidechains is allowed). The ligand is mostly regarded as flexible. Docking can be performed by placing rigid molecules or fragment into protein's active site using different approaches like the *clique-search*, *geometric hashing* or *pose clustering*. In the *clique search* matches are searched to describe the compatible characteristics (shape or interaction pattern) of ligand and protein by means of a distance compatibility graph. This approach is implemented for example in *DOCK* [249]. A *geometric hashing* function is created to describe geometric features like distances in two steps, the preprocessing phase and the recognition phase. This approach is attractive regarding its time-efficacy and the option of partial matching of the ligand in the protein pocket [250,251]. *Pose clustering* is an algorithm based on the matching of triplets of features of the ligand with a triplet of features of the protein. This approach is implemented in the program *LUDI* [252,253].

The first method introduced to treat ligand flexibility was to dock a set of conformers of the same molecule covering the conformational space in an exhaustive way. Unfortunately, different studies [254,255] showed that many conformers are needed (up to few hunderts) to exhaustively sample the ligand conformational space, thereby the demand of the computer time increases.

A more sophisticated approach is the implementation of algorithms that simulate flexible docking, like incremental fragmentation (*FLEXX*) [256,257], molecular simulation (simulated-annealing in *AutoDOCK* [258], molecular dynamics [259]) or Monte-Carlo based methods (*PRODOCK* [260–262], *Yeti* [248,352]). Also docking techniques based on genetic algorithms showed impressive results (*GOLD*) [263,264].

A critical issue that has to be kept in mind when performing docking, is the multi-factorial dependence of the docking results. Aside from docking algorithms and scoring function, binding-site definition and the use of pharmacophore constraints are decisive [265]. Moreover, also the nature of the biological target, the properties of the active site, crystallographic resolution, as well as ligand flexibility and size were found to influence docking reliability [266]. A study by Wu *et al.* [267], clearly showed that a higher docking success (meaning a root square deviation between top ranked docking position and X-ray structure of less than 2Å) could be obtain with a lower number of rotatable bonds. In an other study by Perola *et al.* [268], docking accuracy was related once again to ligand flexibility, but also with the predominant nature of the interaction between protein and ligand, and the degree of solvent exposure of the binding pocket.

1.7.2.2 An insight into the scoring procedures

Scoring of docking poses is still regarded as one of the major challenges in the field of molecular docking. The purpose of the scoring procedure is the identification of the correct binding pose by its lowest energy value, and the ranking of protein-ligand complexes according to their binding affinities [269].

Though much effort has been invested in the development of accurate scoring functions, still none of them can correctly rank every protein-ligand complex today. Two approaches may help to overcome this problem. First, the identification of a good scoring function for the specific case of interest or the application of a consensus-scoring scheme (combination of several scoring functions). Another problem to overcome for a correct scoring arises from the occurrence of induced-fit binding mode of ligands. Compared to the flexibility of small molecules, the tackling of the flexibility issue in macromolecules is still a demanding problem. Hence, the induce-fit protein movements must be regarded as a crucial task to develop better docking and scoring algorithms.

Scoring functions can be divided in empirical scoring functions, scoring functions derived from force fields, and knowledge-based scoring functions. Scoring functions derived from force fields handle the ligand binding prediction with the use of potential energies and sometimes in combination with solvations- and entropy-contributions (*Yeti* [248,352], *Quasar* [270–272,351]). Knowledge-based scoring functions are based on atom pair potentials derived from structural databases. Finally, empirical scoring functions are derived by correlating experimental binding affinities with docking scores through the use of a training set. For exhaustive reviews on the scoring-function problematic see references [273–282].

Recently an interesting study by Schulz-Gasch and Stahl [265] compared the docking results obtained with the programs *FRED*, *GLIDE*, and *FLEXX* in combination with different scoring functions for seven relevant pharmaceutical targets. They divide the scoring functions in "soft" and "hard". A "soft" scoring function is defined as a function not including directional terms, a "hard" one is obviously the contrary. Moreover, they also distinguish between the scoring function used by the docking algorithms (they call it "objective") and for the final ranking (for them the real "scoring"). Although other influences, like binding site definition, are also crucial, they were able to derive a number of general guideline for an easier selection of the best combination of objective- and scoring-function.

In their opinion, for lipophilic binding sites mulitconformer docking with a soft objective and a harder scoring function is particularly suited. If polar groups also play a certain role, incremental construction algorithms, which are always combined with hard objective functions, should be applied. Very polar sites with a dense network of directed interaction require incremental construction and hard scoring functions, or alternatively multiconformer docking in combination with a hard scoring function.

It has, however, to be noted, that particular attention has to be paid when handling large molecules, very polar ones or molecules presenting themselves as tautomers. In fact, practically all scoring-functions overestimate the affinity of large molecule, because of taking unspecific interactions with the protein too much in account [283,284]. In the case of very polar substances, often they are too much taxed by a desolvation- and/or an entropic-term present in the scoring-function equation.

The role of tautomers and of protonation-states in general, has been recently highlighted by Pospisil *et al.* [285]. Thus, molecular-modeling applications usually disregards the last two issues. Obviously just the change of the protonation or tautomeric state of a molecule can dramatically modify its hydrogen-bonds pattern, thereby dramatically influencing the fitting of the ligand in the binding site. Hence, it is easy to understand that the choice of the correct tautomeric or protonation state for the docking experiment and the following evaluation by a scoring function plays a determinant role.

1.7.3 Quantitative structure-activity relationship (QSAR)

Within congeneric series of compounds, quantitative structure-activity relationships (QSAR) correlate affinities of ligands to their binding sites, inhibition constants, rate constants, and other biological activities, either with certain structural features (Free Wilson analysis) or with atomic, group or molecular properties, such as lipophilicity or polarizability (Hansch analysis) [286–298]. Although the relationship between lipophilicity and biological properties like narcotic effect have been known since 150 years, it was only in 1964 that the milestone of QSAR was set through the publication of the Free Wilson method [299] and of the Hansch analysis [300]. Since then, QSAR equations have been used to describe the correlation between biological activity of compounds and some of their physico-chemical properties. Only after a very slow development, the breakthrough for the application of QSAR methods was achieved. In 1988 Cramer *et al.* [301] published the first real 3D QSAR method known as comparative molecular field analysis (CoMFA).

The CoMFA method was developed as a tool to investigate three-dimensional structure-activity relationships. A CoMFA analysis starts with traditional pharmacophore modeling to suggest a bioactive conformation of each molecule and ways to superimpose the molecules under study. This is one of the most critical and difficult steps in a CoMFA study [302,303] and has then later leaded to the development of 4D-QSAR methods (see below). The underlying idea of CoMFA is that differences in a target property like biological activity are often closely related to equivalent changes in the shapes and strengths of the non-covalent interaction field surrounding the molecules. Usually, only two potentials, the steric potential, expressed in a Lennard-Jones function, and the electrostatic potential, expressed in a simple Coulomb function, are used within a CoMFA study. As a result of a CoMFA study, contours map, highlighting regions in space favorable or unfavorable for ligand-receptor interactions of steric or electrostatic nature, are obtained.

Due to the problems associated with the functional form of the Lennard-Jones potential (more precisely the A/r^{12} repulsion term) [304] and the difficulty of a correct parameter setting in most of the CoMFA methods [305,306], Klebe *et al.* [307] have developed a similarity indices-based CoMFA method, which is named

CoMSIA (Comparative Molecular Similarity Indices Analysis). The method uses Gaussian-type functions instead of the traditional CoMFA potentials. Three different indices related to steric, electrostatic and hydrophobic potential are computed. The advantage of this method lies in the function used to describe the molecules studied, as well as in the resulting contours maps. In fact, these maps are easier to interpret than the ones obtained with the CoMFA approach. Moreover the CoMSIA procedure avoids the cutoff values used in CoMFA to prevent the potential functions from assuming unacceptably high values (i.e. unfavorable). For a more detailed description of the CoMSIA procedure refers to [308–310].

A further refinement of the original CoMFA technique has been achieved by introducing the concept of variable selection and reduction [311,312]. Hence, the large number of variables in the descriptor matrix is one of the problems of the CoMFA method. This could be partially solved by the introduction of a statistical procedure called GOLPE (General Optimal PLS Equation) developed by Barone *et al.* [311].

Several other 3D QSAR approaches have been developed during the last few years and some of them differs from the CoMFA method by not being based on properties calculated within a lattice. In *GERM* [313], *COMPASS* [314] and *Quasar* [270–272,351], properties are calculated for discrete locations in the space at or near the union surface of the active ligands. The "receptor surface" thus generated is intended to simulate the macromolecular binding site. In the past, these methods had two major drawbacks associated with the atomistic and receptor-surface based on averaged receptor entities: the adaptation of the shape of the binding site by means of induced fit and hydrogen-bond adaptation. However, these problems have been approached in the meantime. Vedani *et al.* [270–272] developed so-called multidimensional QSAR methods (4D–, 5D–, and 6D–QSAR). The 4D extension refers to the possibility to allow for more than one conformer per molecule or to handle different orientations, protonation or tautomeric states of a compound at the same time, letting thereby the genetic algorithm select the most suited conformation of each molecule to obtain the best possible model. A further extension (5D) was introduced in 2002, when the possibility to simultaneously handle different induced-fit scenarios was implemented. Finally, the necessity to account also for different solvation models was recognized and implemented in the software *Quasar* [270–272,351], leading to the actual 6D version. In 2004 the same group developed also another multi-dimensional (4D) QSAR approach, implemented in the software *Raptor* [315, 316,335]. In this case, the fourth dimension is used to handle anisotropic induced-fit of the protein in the presence of different ligands. The binding site is therefore, represented by two layers. The inlying layer maps the fields, which a substrate would feel, if it were to fit snugly into the binding pocket. Another compound, that features additional groups that reach deeper into the protein, may experience different fields as a consequence of induced fit. This can be modeled by a second, outer layer. For major details refer to [315,316] and to *Chapter* 3.1.2.10.

1.7.4 *De novo* Design

De novo design has the difficult goal to design novel molecular structures with the desired pharmacological properties starting practically from scratch. During the last two decades a great variety of *de novo* design software has become available. In *de novo* design, the following questions are addressed: (i) how to assemble candidate compounds; (ii) how to evaluate their potential quality (pharmaco-dynamic and pharmaco-kinetic properties); and (iii) how to sample the search space effectively for new ligands. A summary of the different approaches is given in *Table* 4.

Table 4: Selected *de novo* design programs with their basic properties in chronological order (adapted from [317]).

Name (year)	Building blocks		Primary target constraints		Search strategies					Structure sampling					Scoring function
	At	**Fr**	**Rc**	**Li**	**DFS**	**BFS**	**Rnd**	**MC**	**EA**	**Gr**	**Lk**	**Lat**	**MD**	**Sto**	
3D Skeletons (1990)		X	X		X					X					Steric constraints and hydrogen bonds
Diamond Lattice (1990)	X		X		X							X			Steric constraints and hydrogen bonds
Legend (1991)	X		X				X			X					Force field
LUDI (1992)		X	X			X				X	X				Empirical scoring function
NEWLEAD (1993)	X	X	X			X					X				Steric constraints
SPLICE (1993)		X	X			X					X				Pharmacophore and steric constraints
CONCEPTS (1993)	X		X					X					X		Empirical scoring function
Growmol/Allegrow (1994)	X	X	X					X		X					Simple empirical scoring function
MCDNLG (1995)	X		X					X						X	Potential energy
Chemical Genesis (1995)		X	X	X					X					X	Combined score and shape, grid-based and scalar constraints
PRO_LIGAND (1995)		X	X	X	X					X	X				Empirical scoring function
SmoG (1996)		X						X		X					Knowledge-based scoring function
SkelGen (1997)	X	X	X					X						X	Geometric, connectivity and chemical constraints
Topas (2000)	X		X					X						X	Molecular similarity based on topological pharmacophore and substructure fingerprints
CoG (2004)	X	X		X				X						X	Molecular similarity based on fingerprint descriptors

At: Atoms; **Fr**: Fragments; **Rc**: Receptor; **Li**: Ligand; **DFS**; Depth-first search; **BFS**: Breadth-first search; **Rnd**: Random; **MC**; Monte-Carlo sampling with Metropolis criterion; **EA**: Evolutionary algorithms; **Gr**: Grow; **Lk**: Link; **Lat**: Lattice; **MD**: molecular dynamics simulation; **Sto**: Stochastic

In this chapter, only protein-based methods will be briefly discussed. The first step of most of the protein-based *de novo* design software is to generate the so-called primary target constrains by generating a grid of points in the binding site, and computing the interaction energies by placing different probe atoms or fragment at each grid position. An appropriate selection of probes and of the grid resolution, leads to the identification of the key interaction sites. This first step

results in maps highlighting the pinpoints favorable for positioning a particular atom (H-bond donor/acceptor, hydrophobic atom) of the searched ligand. After that, the search for a new ligand can start. Several concepts have been developed for the structure sampling: *linking, growing, lattice-based sampling, random-structure mutation, transition driven by molecular dynamics simulations*, and *graph-based sampling*.

Linking and *growing* have emerged as the most promising (*Figure* 30). The *linking* approach starts with the placement of building blocks at key interaction sites of the receptor. These building blocks are either positioned by the program itself or by the user [317]. The positioned building blocks are then automatically connected to each other by so-called linkers to yield a complete molecule that satisfies all key interaction sites. Linkers are selected from a library (e.g. amide, sulfonamide, indol rings, benzene,…) with the objective of forming favorable interaction with the target protein.

In contrast, the *growing* procedure starts with a single building block at one of the key interaction sites of the receptor. This starting point is set either by the program or the user. The structure is then grown from the initial building block in an attempt to provide suitable interactions for both the key interaction sites of the receptor and region of the receptor between two key interaction sites. The main problem for the *de novo* design algorithms is how tackle the combinatorial explosion. In fact, it is easy to imagine how many different structures can be generated also in a small binding site by using only few fragments. Therefore, methods for a fast evaluation of the partial solutions (e.g. solutions satisfying only two key interactions sites) and thereby drive the search were needed.
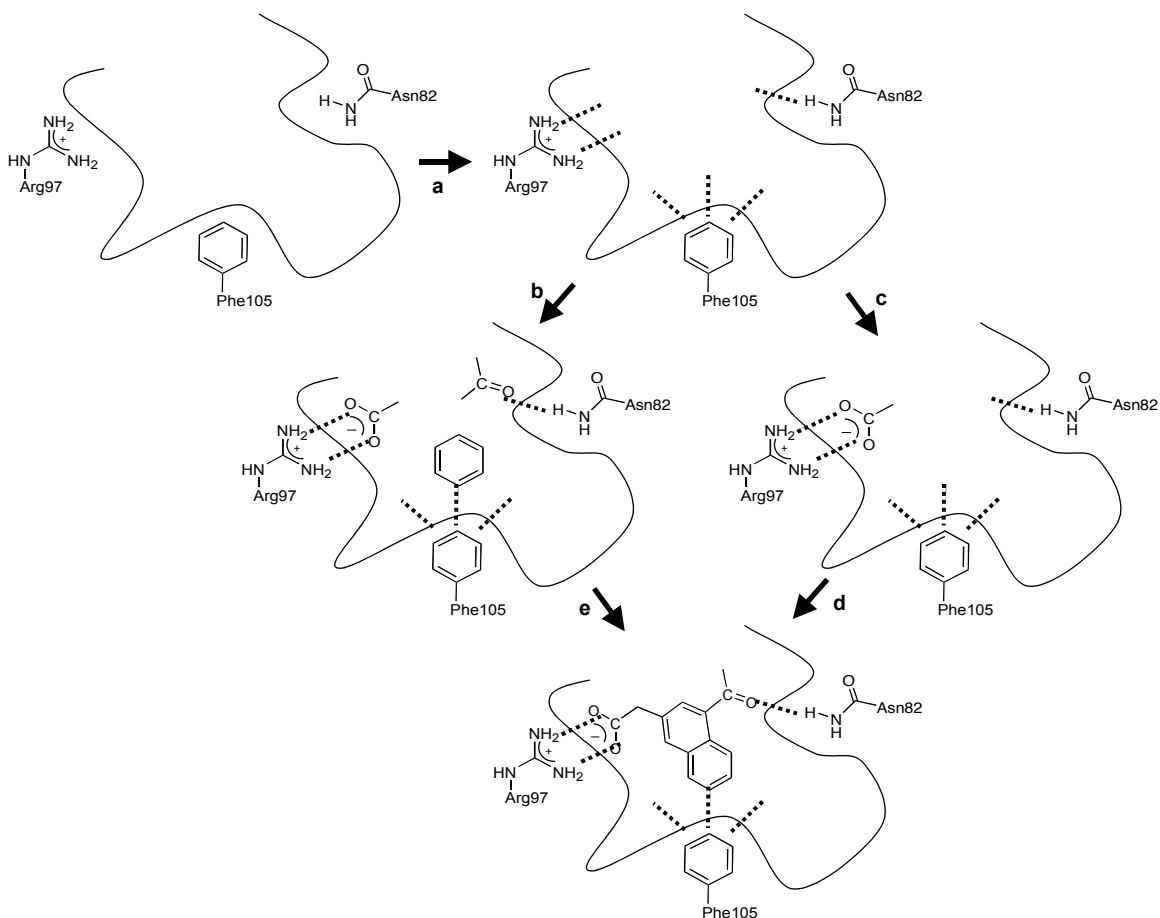
Figure 30: Principle of structure-based ligand design (adapted from [317]). **a**) determine the interaction points; **b**) place the fragments; **c**) place the first fragment; **d**) *grow* strategy; **e**) *link* strategy.

Four different approaches to tackle this problem were developed: the *breadth-first* strategy, the *depth-first* strategy, *Monte-Carlo methods*, and the use of *evolutionary algorithms*. A *breadth-first* strategy (*Figure* 31) retains all partial solutions at one level of the search space graph and explores, sequentially, other levels until each of these paths reaches an end state. During a *breadth-first* search all nodes are systematically examined so that identifying the optimal solution is guaranteed. This exhaustive procedure is very time consuming and therefore only implemented in software using the *linking* sampling method of already positioned fragments. Thereby it deals with a relatively small combinatorial space (only different linkers are available), making this approach applicable. In contrast, the *depth-first* strategy (*Figure* 31) retains only one of a variety of possible partial solutions at each of the search space graph, until an end state is reached. The choice of the single retained partial solution can be guided by its score, by chance or by a combination of both. The third approach is bases on random sampling (*Monte-Carlo*) usually controlled with the Metropolis

criterion. In this procedure, after each structure-modification step, the change is evaluated to decide whether it is accepted or rejected. If the modification results in an energetically more favorable candidate compound, it is immediately accepted. If, on the other hand, the modification yields to a candidate compound with higher energy, it can still be accepted with a probability that is based on the scoring function difference between the modified and unmodified structure and a random number. This approach is for instance implemented in *AlleGrow* [318,319]*,* the *de novo* design software used for this work and will be presented in detail in *Chapter* 3.1.2.1. The fourth approach to tackle the problem of combinatorial explosion is to make use of *evolutionary algorithms*. Candidate compounds are represented as individuals of a population, their suitability is then evaluated by a so-called fitness function that determines which structures are chosen as parents for the next generation.

Once the target number of structures required has been generated, the candidate compounds are usually sorted by a scoring function and visually inspected by the user, which identifies the most promising ones for other runs of *de novo* design or synthesis. The newest software, as for example *Skelgen* [320,321], incorporates not only typical scoring functions to rank the obtained compounds, but also the possibility to rank them by their physico-chemical properties and/or by their chemical feasibility, thereby simplifying the choice for the user.
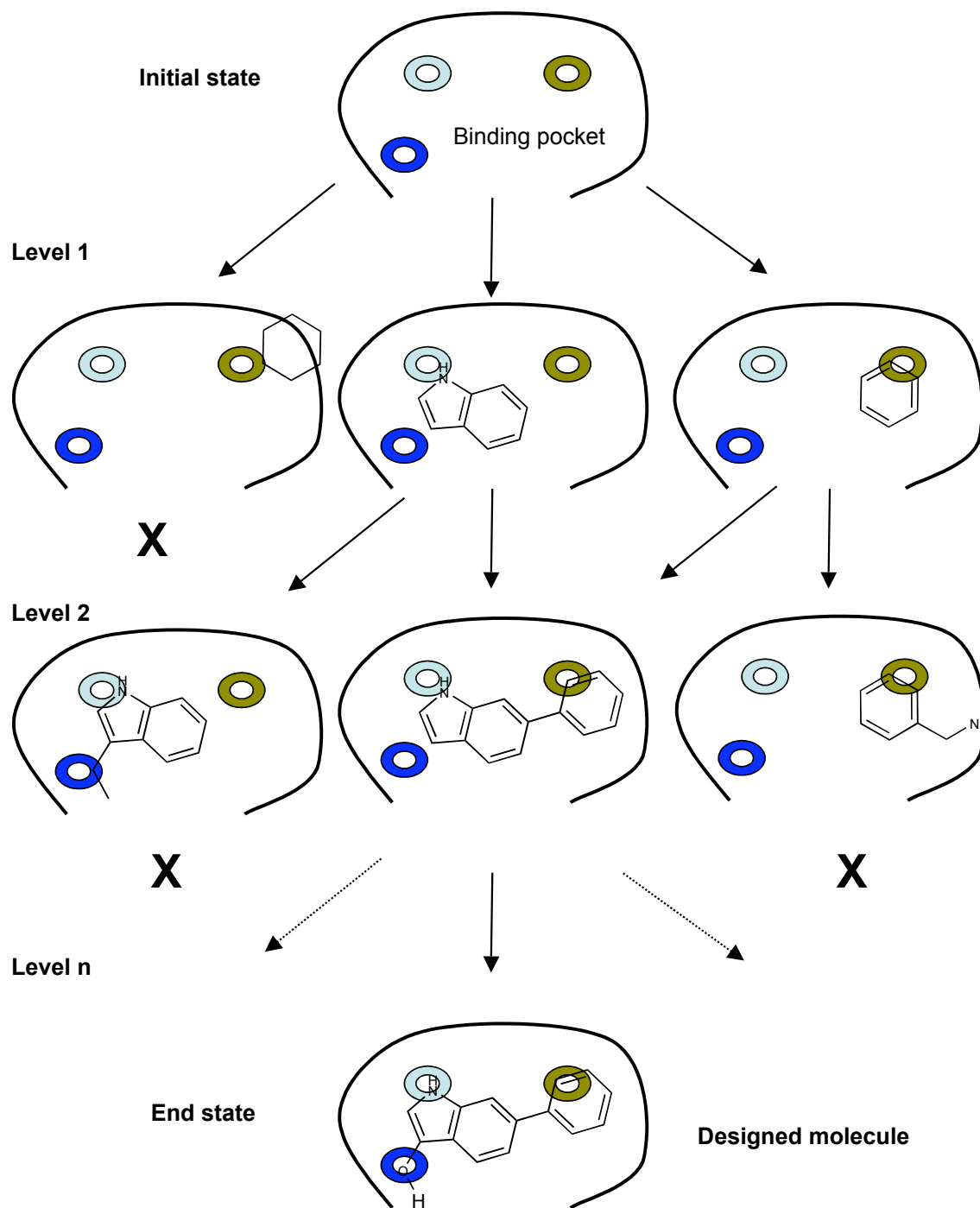
Figure 31: Tree model of search space exploration by an automated structure generation method. A binding pocket and predicted key interaction points are given as an input (entry state). In this example, key interaction points are pharmacophore points at which light blue circles indicate sites for a ligand hydrogen-bond donor, dark blue for hydrogen-bond acceptor, and orange for lipophilic ligand parts. Candidate compounds are assembled by *growing*. In the example some partial structure are rejected due to bondary violation (Level 1 and 2) or due to

mismatching interaction type (Level 2). Two partial solution emerge from Level 1. A *breadth-first* search would follow both nodes in the search graph simultaneously, whereas the three shown here selects a single node for expansion (*depth-first search*). The picture is adapted from [317].

2 Aim of the thesis

E-selectin, a member of a family of cell-adhesion proteins, which plays a crucial role in many physiological processes and diseases [1], and in particular, in the early phases of the inflammatory response. Its role is to promote the tethering and the rolling of leukocytes along the endothelial surface [2]. These steps are then followed by integrin-mediated firm adhesion and final transendothelial migration. Henceforth, control of the leukocyte-endothelial cell adhesion process may be useful in cases, where excess recruitment of leukocytes can contribute to acute or chronic diseases such as stroke, reperfusion injury, psoriasis or rheumatoid arthritis [3]. In this work, our efforts to develop *in silico*-based protocols to study the interaction between E-selectin and its ligands, will be presented.

First, different protocols, established as a part of this work to accomplish this task, will be illustrated and discussed. Particular importance will be given to the development of protocols for the analysis of the conformational preferences of E-selectin antagonists, and to the generation of QSAR-models using *Quasar* [270–272,351] and *Raptor* [315,316,335]. Hence, conformational issues seem to play a fundamental role to determine the affinity of the ligands towards the target protein (cf. *Chapters* 1.5.2, 1.5.3). Moreover, the derivation of QSAR-models will permit to have at disposition an efficient tool for a preliminary *in silico* affinity estimation of new designed ligands.

Second, the application of the developed protocols to the design of new E-selectin antagonists will be debated.

3 Material and Methods

3.1 Material

During this thesis various computer systems and software were used. Detailed are given in *Table* 5.

Table 5: Overview of the computer, operating systems and software used.

Computer Model	Processor	Memory	Operating System	Modeling Software
Macintosh PowerBook G4	PowerPC G4 (2.1) 667MHz	512MB	OSX 10.2.1-10.3.8	*Yeti* 6.0–6.6 *Excel*
Macintosh Power Mac G5	Dual PowerPC G5 (2.2) 1.8GHz	1GB	OSX 10.3.2	*Yeti* 6.0–6.6 *Raptor* 1.2 *Quasar* 5.0 *AMBER* 7.0 *Pymol* 0.98
Macintosh Power Mac G4	PowerPC G4 (2.1) 400MHz	384MB	OS 9	*Yeti* 6.0
Dell Precision^{TM} Workstation 650	Dual Intel® Xeon^{TM} 2.8GHz	1GB	Red Hat 8.0	*AMBER* 7.0 *MOPAC* 6.0 *VMD* 1.8.2 *Plotamber* 0.55
Cronos Linux Cluster (for more details see Pic XX)	76x Intel® Xeon^{TM} 2.4GHz	2GB for each node	Red Hat 8.0	*AMBER* 7.0
Silicon Graphics Workstation Octane	Dual 250 MHz MIPS R 10000	1.2GB	IRIX 6.5	*MacroModel* 5.0 *Rfitmm* *PrGen* 2.1 *AMSOL* 5.4.1 *AMBER* 7.0 *QXP* *Allegrow*
Dell Precision^{TM} Workstation 530	Dual Intel® Xeon^{TM} 2.4GHz	1GB	Windows XP	*Samoa* 0.98

3.1.1 Software

In this section all software used in this work will be introduced. The detailed protocols applied, are instead presented in *Chapter* 3.2.

3.1.1.1 Allegrow/QXP

Allegrow [318,319] allows to generate hypothetical ligands with topological and chemical complementarity to the three-dimensional structure of a host target protein, starting from a growing point of a known ligand or of a fragment previously docked in the binding site or from an experimentally determined structure. The molecules are generated in the host-binding site fragment by fragment, atom or functional groups. At each step the position and type of atom to be added is randomly selected from a set of possible fragments, which are consistent with internal bond lengths and bond angle requirements as well as the spatial and chemical properties of the binding site. The selection is achieved using Boltzmann statistics to enhance the acceptance towards atoms, which can form favorable interactions with the binding site. Atoms, which are less complementary to the binding site will also be chosen but with a low probability. This allows ligand structures to be formed in which groups of atoms with favorable interactions can be connected by atoms, which are less complementary, driving however the search towards overall higher complementary ligands. Another important feature of *Allegrow* [318,319] is its ability to connect a newly generate atom to a previously grown atom in the growing structure. In that way, rings can be obtained by connecting closely previously positioned non-bonded atoms. Conformationally restricted molecules may thereby be obtained, having a favorable entropic effect on the binding energy. The Boltzmann-driven process ensures that a highly diverse set of molecules is generated unbiased by previously generated candidate structures. After this generation step, the obtained molecules are analyzed automatically to avoid duplicates and structure with too high conformational strain energy. Those are eliminated and the remained solutions minimized with the QXP force field [319], and finally sorted by a means of a scoring function [318,319]. Technical details are given in [318,319].

3.1.1.2 AMBER 7.0

AMBER 7.0 [322] is one of the worldwide most used packages and includes a suite of programs (MM, MD, FEP, GIBBS, ptraj,…). In this work, it was exclusively used to perform MD simulations in explicit solvent [323]. The programs used include: *antechamber* (ligand preparation) [324], *Leap* (ligand and target preparation) [322], *sander* (minimization and molecular dynamic simulations) [322], and *ptraj* (analysis) [325]. The description of the protocols used to investigate the conformational preferences of E-selectin antagonists and the stability of E-selectin complexes are described in the *Chapters* 3.2.3.2 and 3.2.6.

3.1.1.3 AMSOL 5.4.1

AMSOL 5.4.1 [326] permitted to calculate CM-1 charges [327] and solvation energies. The calculations were performed at the *Biographics Laboratory 3R* [328].

3.1.1.4 Excel

Excel was used to analyze the results of the MD simulations and of the conformational studies.

3.1.1.5 MacroModel 6.5

MacroModel 6.5 [169] was used to generate the 3D-structure of all E-selectin antagonists studied, to perform conformational searches (CS) and Jumping between Wells (JBW) simulations (cf. *Chapter* 3.2).

3.1.1.6 MOPAC 6.0

MOPAC 6.0 [329] was used to calculate AM1-BCC charges [330] or MNDO-ESP charges [331]. The last were kindly generated at the *Biografics Laboratories 3R* [328].

3.1.1.7 Plotamber 0.55

This program [332] permits to extract the information contained in an AMBER-output file into a text-file, which can then be imported into Excel and used to plot for instance the course of the energy of the system during an MD simulation.

3.1.1.8 PrGen 2.1

PrGen 2.1 (Pseudoreceptor Generator) [333] was mostly used to prepare the compounds for receptor modeling studies or docking studies. In particular, for performing their alignment and to assign the entropy (TΔS) change upon ligand binding of the individual ligands. The preparation included also renaming of the ligands, assigning force-field types and adding the experimental binding energies

(IC_{50}). Additionally, the software gives the possibility to prepare *runfiles* for *AMSOL-* [326,327] or *MOPAC*-calculations [329–331]. *PrGen* [333] dealt also as "converting platform". The software was often applied to convert files in *MacroModel* format into files in PDB- or PDB-extended-format.

3.1.1.9 Pymol 0.98

Some pictures within this thesis, were generated with *Pymol* [334].

3.1.1.10 Raptor 1.2

Raptor [315,316,335] was applied to generate models for the prediction of the relative free binding energy of ligands. It is based on a multi-dimensional quantitative structure-activity relationship. In addition to the usual 3D structure-activity relationship, a fourth dimension accomplishing for the induced-fit phenomena is introduced. Hence, induced fit is explicitly and anisotropically allowed by a dual-shell representation of the receptor surrogate. The scoring function treats solvation effects implicitly and is therefore independent from partial-charge models. Moreover, *Raptor* [315,316,335] offers the possibility to define a threshold for the weak binders. This function allows gaining valuable information from the weak binders ligand molecules about the binding site of the protein without punishing the prediction of weaker binding than experimentally measured. A *scramble test* verifies the biological significance of the generated model [338]. More information about the software can be found in [315,316,335].

3.1.1.11 Rfitmm

Rfitmm [336] is a in-house developed software that permits to superimpose the n best conformers of a structure obtained from a conformational-search study performed with *MacroModel* [169] and sort them energetically.

3.1.1.12 Samoa 0.98

Samoa [337] is a piece of software developed at the Institute of Molecular Pharmacy by Samuel Schmid. A detailed description of the program can be found in his thesis [337]. In this work, it was mainly used for analytical purposes (Φ/Ψ-plots, ^{3}J-coupling calculations out of MD simulations, tracking of distance or angles over the simulation time,…) and to automatically generate the command files for the JBW/SD-simulations (cf. *Chapter* 3.2.3.1).

3.1.1.13 Quasar 5.0

The software was used to generate a receptor model for E-selectin. *Quasar* [270–272,351] – a receptor-modeling concept developed at the *Biographics Laboratory 3R* [328] – is based on 6D–QSAR combined with a directional force field and explicitly allows for the simulation of induced fit. In generates a family of quasi-atomistic receptor-surface models that are optimized by means of a genetic algorithm [338, 339]. This evolution is driven by the lack of fit (*LoF)* function. Based on this function, two individuals of the population with the highest *LoF* – considering the number of properties mapped onto the surface, the similarity of the models and the specificity of the conformer selection – are discarded at the end of every crossover to improve the cross-validated r^2 ($\rightarrow$ q^2) of the training set. The hypothetical receptor site is characterized by a three dimensional surface which surround the ligand molecules at *van der Waals* distance and which is populated with atomistic properties mapped onto it. The topology of this surface mimics the three-dimensional shape of the binding site: the mapped properties represent other information of interest, such as hydrophobicity, electrostatic potential, and hydrogen-bonding propensity. In *Quasar* [270–272,351], the fourth dimension ($\rightarrow$ 4D–QSAR) refers to the possibility of representing each ligand molecule as an ensemble of conformations, orientations, and protonation states, thereby reducing the bias in identifying the bioactive conformer. Within this ensemble, the contribution of an individual entity to the total energy is determined by a normalized Boltzmann weight. In contrast to other concepts in the field, *Quasar* [270–272,351] allows for the explicit simulation of induced fit. The algorithm allows simultaneously evaluating of different induced-fit scenarios ($\rightarrow$ 5D–QSAR). Presently, those include a linear fit, an adaptation based on the total interaction energy as well as four protocols reflecting the steric, electrostatic, hydrogen bond and lipophilic potential, respectively. Solvation phenomena – such as ligand desolvation, solvent stripping or shielding effects – may jeopardize an otherwise robust simulation as they directly (solvation term) or indirectly (induced-fit component) contribute to the calculated binding affinity.

In *Quasar* [270–272,351], this quantity based on a directional force field [248,341] is determined as follow:

$$E_{binding} = E_{(force\ field)} + E_{polarization} - E_{(ligand\ desolvation)} - T\Delta S_{binding} - \Delta E_{(internal\ strain)} - E_{(induced\ fit)}$$

(**Eq. 4**)

Free energies of ligand binding ΔG_{pred}, are then predicted by means of a linear regression between ΔG_{exp} and E_{bdg} using the molecules of the training set:

$$\Delta G_{pred} = |a| * E_{binding} + b \quad \textbf{(Eq. 5)}$$

The sixth dimension ($\rightarrow$ 6D–QSAR) allows for the simultaneously consideration of different solvation models (e.g. degree of desolvation, contribution of different protonation states). In addition, it may be used to scale the influence of entropy to ligand binding and internal strain in an unbiased fashion.

The correlation coefficients are q^2 (cross validated r^2) for the training and p^2 (predictive r^2) for the test set. Ideally they reach the value of one. Another parameter describing the quality is the root mean square deviation (*rms*), which should approach zero. The quality of the prediction is validated using an external set of compounds ($\rightarrow$ test-set). For this purpose experimentally measured binding affinity of the ligands are required. Best predictions of the binding affinities were obtained with a large set of molecules in combination with a ratio of 3:1 for training- and test-set.

To verify the sensitivity towards the biological data, the experimental binding affinities are scrambled and the simulation is repeated otherwise identical. A *scramble test* with comparable results to the predicting test indicates an apparent good model worthless [338].

The program allows for a functional-group analysis that identifies the individual contribution of common functional groups of the compounds to the binding affinity — a useful tool for further developing of high-affine ligands.

As any QSAR study, *Quasar* [270–272,351] is typically only applicable within a series of related structures, which implies that QSAR models are restricted to a specific field of research.

3.1.1.14 VMD 1.8.2

VMD (Visual Molecular Dynamics) [341,342] is a molecular graphics program designed for the interactive visualization and analysis protein-ligand complexes. The software is able to read in different formats of molecular trajectories and to visualize MD simulations. It gives thereby the possibility to generate movies or still pictures that are useful for the characterization of the ligand and/or protein motion. Moreover it can be used to monitor distances, angles or dihedrals over the time.

3.1.1.15 Yeti 6.0-6.6

Yeti [248,352] is a molecular-mechanics program developed by Vedani and Huhta to optimize metalloproteins complexed with small molecules. Its main purpose is the minimization of molecules and complexes. In this work, the Monte-

Carlo minimization protocol with Boltzmann sampling was used to identify the various conformations of the ligand molecules at the binding site and to obtain thereby the most feasible docking mode. The energetically most feasible pose was then usually submitted to MD or used for receptor-modeling studies.

3.2 Methods

3.2.1 Data Flow, Strategy

The data flow and the strategy adopted during this thesis is presented in *Figure* 32.
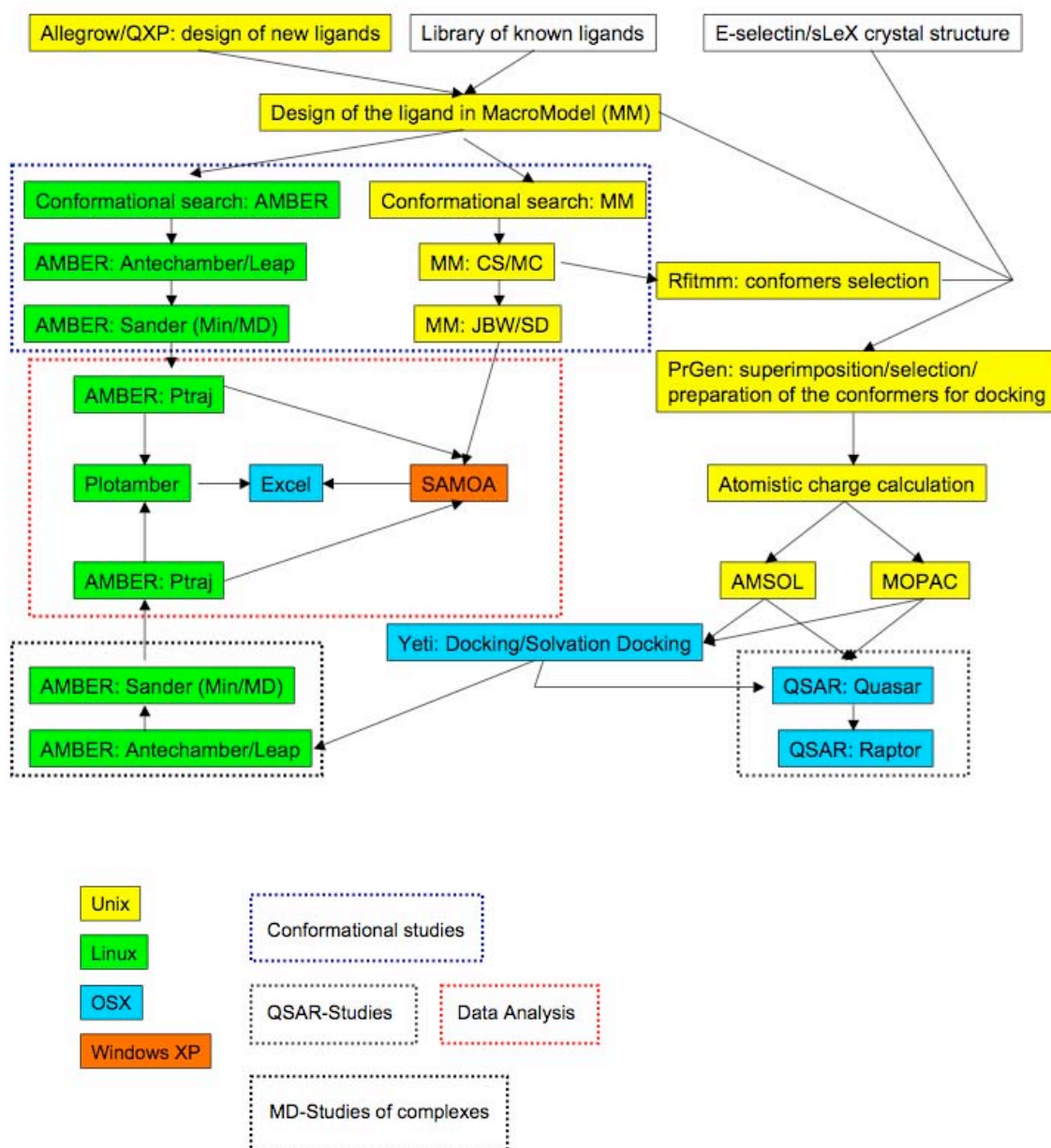
Figure 32: Schematic representation of the strategy adopted in this work and of data flow.

3.2.2 Ligand design with MacroModel

The three-dimensional structures of all the ligands studied, was generated using *MacroModel* 6.5 [169] and optimized in aqueous solution [171] on the basis of the improved AMBER 4.0 force field for carbohydrates published by Still [170] and partially modified by Kolb and Ernst [161,168]. In principle all structures should be submitted to an exhaustive conformational search (CS) but that would have been too time consuming. It was therefore decided to execute such a protocol only with selected molecules (cf. *Appendix* 1) representing thereby most of the structural variety of the E-selectin antagonists possible. All other structures were generated taking advantage of structural similarity, by using the output of those CS-studies as a starting point for the new molecule instead. After the necessary small modifications, the new compound was optimized keeping thereby the common scaffold rigid.

3.2.3 Conformational analysis

3.2.3.1 Conformational analysis using MacroModel

The method is based on the "Jumping between Wells/stochastic Dynamics" (MC(JBW)/SD) algorithms [161,168] and the systematic pseudo-Monte-Carlo (SUMM, **s**ystematic, **u**nbounded **m**ultiple **m**inimum search) simulation technique [231] . A graphic summary of the method is presented in *Figure* 33.
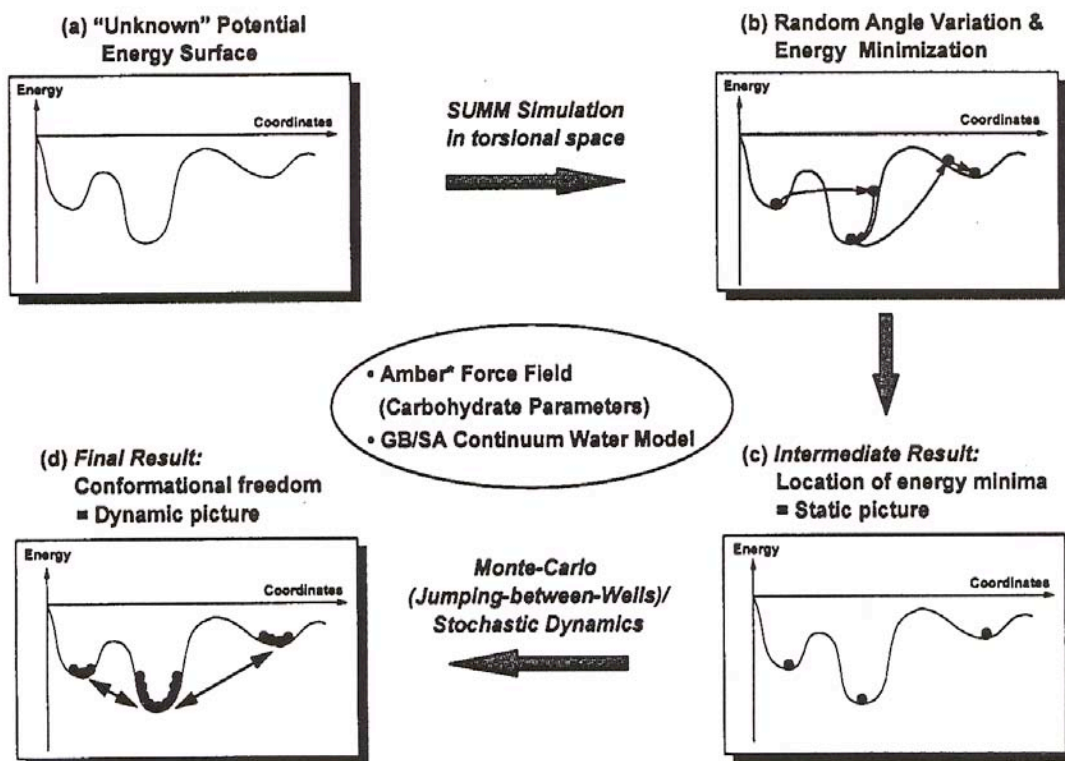
Figure 33: Analysis of the potential energy surface with the (MC(JBW)/SD)-technique.

First, the locations of the most relevant energy minima (conformations) of a compound were determined in an internal coordinate systemic pseudo-Monte-Carlo SUMM simulation [231]. The number of steps performed, were of 2000 for each free-rotatable non-terminal bond. At the end of this pseudo-Monte-Carlo step, only structures within a 20 kJ/mol from the energy of the global minimum were retained. The shape of the potential energy surface was then probed in a subsequent MC(JBW)/SD simulation which used the information obtained in the SUMM analysis. Thus, a Boltzmannn-weighted ensemble of states was generated in a 10ns MC(JBW)/SD simulation by jumping between different energy wells, i.e. the energetically best 100 conformations found in the preceding SUMM analysis, and performing stochastic dynamics simulations within each well. All calculations were performed with an AMBER force field augmented by paramters for carbohydrates [161,168,170] and in conjunction with the GB/SA continuum water model [171]. As a result, the distribution of the conformers of the structure of interest in respect to the whole conformational space was obtained. The exact setting of this kind of simulations can be seen in the commented command-files of the two steps presented in the *Appendix* 2. The command file for the JBW/SD-step was automatically generated by *SAMOA* [337].

3.2.3.1 Conformational analysis using AMBER

The *AMBER*-based protocol used to study the conformational behavior of the E-selectin antagonists is based on the methodology published by Weimar and Woods in the chapter "Combining NMR and Simulation Methods in Oligosaccharide Conformational Analysis" of the book "NMR Spectroscopy of Glycoconjugates" [343]. The flow chart of this method is presented in *Figure* 34.

THE AMBER-BASED PROTOCOL
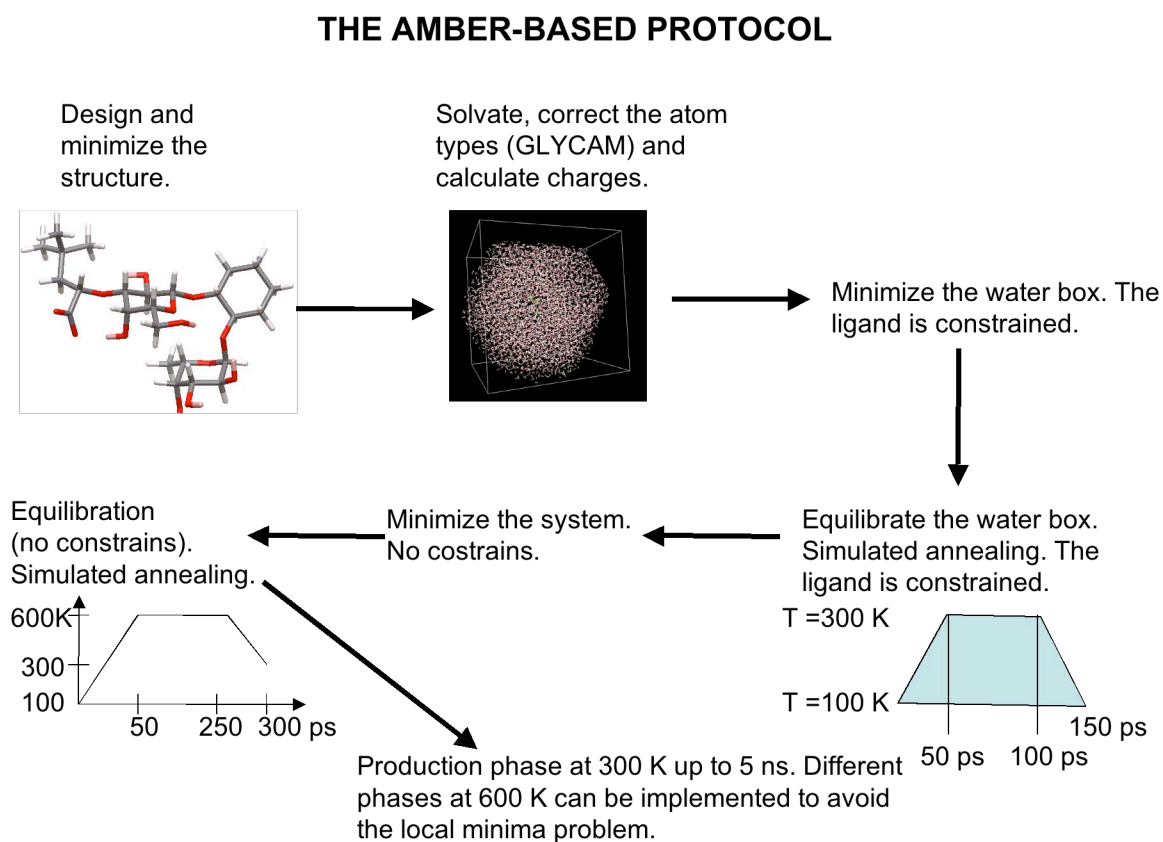
Figure 34: Flow chart of the AMBER-based protocol. This procedure bases on [343].

Preparation of the ligands

As a starting point of my procedure, a *MacroModel*-minimized structure of an E-selectin antagonist was always used. As a first step, the structure of the compound of interest was converted to the PDB-format using *PrGen* [333]. Further, the atom-type of the molecule were manually altered to comply with the *AMBER*-convention, the structure processed with *Antechamber* [324], and atomistic partial charges (AM1-BCC) [330] calculated with *MOPAC 6.0* [329]. At this stage, the atom-types of the glycosidic oxygen atoms, of the oxygen atoms

of $^1C_4$ chair systems, and of the anomeric carbon atoms were changed to fulfill the GLYCAM 2000 [344] requirements. The obtained structure was then read in *Leap*, where it was checked for consistency. In case of missing force-field parameters, those were added to the GLYCAM-parameter file. A list of the parameters added to the original GLYCAM-parameter file can be found in the *Appendix* 3. When all the needed parameters were available, the ligand was solvated [345], and if appropriate counter-ions added. All the studies were performed with the ligand placed into a 20 Å broad pre-equilibrated box of TIP3P water molecule. The topology- and the coordinate-files of the solvated ligand were saved and used as an input of "real" investigation protocol.

Minimizations and MD (Figure 34)

First, an AMBER restart file was generated (cf. *Appendix* 4: getrst.in). This file served as a reference structure during the first minimization step (cf. *Appendix* 5: minres.in), where only the position of the pre-equilibrated water molecules was optimized. As a second step, a 150 ps simulated annealing protocol (cf. *Appendix* 6: annealingNEW.in), involving once again only the solvent was applied. In this step, the temperature of the system was raised during 50 ps from 100 K to 300 K and then, after 50 ps at this temperature, cooled back to 100 K. The obtained system, formed by the molecule of interest, the 20 Å water box, and the counter ion, was then further minimized without this time any constraints (cf. *Appendix* 7: minAN.in) and finally submitted to a long simulated annealing protocol (cf. *Appendix* 8: ANallinone.in). This final step consisted of first period of 50 ps, where the temperature of the system was raised from 100 K to 600 K, a second period lasting 200 ps at 600 K, a third one where the system was cooled down to 300 K followed by the first sampling phase at 300 K lasting one nanosecond, a new warming phase to 600 K of 50 ps, a second period of 200 ps at this high temperature, the second short (50 ps) cooling phase and by the final sampling period lasting again one nanosecond. In some cases the protocol was performed over only 1.3 instead of 2.6 ns.

3.2.4 Preparation steps for the Docking-, MD-, and QSAR-studies

3.2.4.1 Protein preparation for the Docking-, MD- and QSAR-studies

The protein structure co-crystallized with sLe^x (RCSB Protein Data Bank [346] code: 1G1T, resolution: 1.5 Å) solved by Somers *et al.* [166] was retrieved from the RCSB Protein Data Bank [346] and intensively investigated (B-Factors, protonation states of histidines, analysis of the electron-density map,…). The file was then read in *PrGen* [333] where:

- The crystallographic water molecules were deleted

- The polar hydrogen atom added.

- The complex subsequently saved in pdb_extended format (the atomic charges [327] and the solvation energy for the ligand where computed with *AMSOL* [326]) and handle over to the software *Yeti* [248,352].

Here, a first minimization of all rotatable polar H-atoms, followed by a full minimization of the complex was performed. This relaxed structure was then used as reference for all the following experiments.

3.2.4.2 Ligand preparation for the Docking-, MD- and QSAR-studies

The steps needed to prepare the ligands for docking-, MD-, or QSAR-studies that will be presented here, can be graphically followed on *Figure* 32.

The minimized structures in MacroModel-format were read in *PrGen* [333] and aligned with respect to the experimental conformation of sLe^x bound to E-selectin [166] after the minimization described above (cf. *Chapter* 3.2.4.1), using the three O-glycosidic atoms as reference points. The superposition was further improved by manually adjusting the hydroxyl groups essential for binding (Gal–6OH, Gal–4OH, Fuc–4OH, Fuc–3OH) to obtain a common hydrogen-bonding pattern. To simplify the analysis of the computational experiments the ligands were renamed by applying a three-letter code (cf. *Appendix* 9). Next *PrGen* [333] was then used to calculated the entropy (TΔS) penalty for each ligand, to assign the atom types and to read in the individual experimental binding energies of every ligand. Ligand files containing ester and/or nitro groups had to be manually assigned special force-field types. In fact, it is known [347], that these atoms behave particularly with respect to hydrogen-bond formation. Therefore, the force-field atom type of the ester oxygen had to be changed to OX (from OE) and to ON (from O2) for both the oxygen atoms of the nitro-group. Further, *runfiles* for atomic charge calculation with AMSOL [326,327] and MOPAC [329,331] were

prepared. The following steps were the charge calculation performed with AMSOL [326,327] or MOPAC [329,331] respectively, and the computation of the solvation energies done by AMSOL [326,327]. Finally, the ligand files containing coordinates, charges and energy terms such as solvation, entropy and experimental binding energies was saved and then used for the docking- and the QSAR-studies. In contrast, file for MD-studies were saved in standard format (PDB).

The experimental binding energies, which are necessary for any QSAR study were calculated according to *Equation* 6 using the IC_{50} data available in house (cf. *Appendix* 9). All the concentrations are measured in mM. If no exact IC_{50} was specified, the value was set to 20 mM for >10 and to 7.5 mM for >5. In the case of a multiple-entry of the ligand in the database, the averaged IC_{50} was used.

$$\Delta G = R \bullet T \bullet ln IC_{50} \quad \textbf{(Eq. 6)}$$

3.2.5 The docking protocol

3.2.5.1 General procedure

Before running any minimization, the force-field parameter database had to be augmented to include all specific parameters for the studied ligands – most particularly, 1–4 (torsional) interactions (cf. *Appendix* 10). Next, the ligands were manually docked into the binding site of E-selectin and minimized using constraints (corresponding to those in the experimental structures [166]) specified in the *Appendix* 11 to get an ideal starting position for the MC-minimization. The minimization of the complex was performed over three steps: first, only the ligand was minimized (the protein was kept rigid, zone 0.0 Å), then the ligand and an eight-angstroms zone surrounding it were optimized and finally the whole complex was relaxed. The obtained structure was then used as a starting point for the MC-minimization with Boltzmann sapling, which was carried out without constraints. The Fuc–4OH was chosen as the center of rotation for the MC-sampling to avoid unreasonable starting structures. This atom was therefore used as the reference to position and orient the compound during each Monte-Carlo round assigning new starting coordinates. The weighting for the energy components hydrogen-bond and metal-interaction was set to 1.2, which means that they are considered more in the selection process. This value resulted to be the more suitable from a series of pre-experiment (data not reported here). The criteria to be fulfilled for a structure being accepted for the Boltzmann-sampling step was set to a maximal energy difference of 0.6 kcal/mol between the starting structure and the actual lowest energy solution. If none of the starting structures could fulfill the criteria, the software automatically save the structure with the lowest starting energy. The sampled structures were written out

at a frequency of ten rounds. Thus, from a MC-minimization of 2,000 rounds 200 different local minima conformations were saved. More details on the setting of MC-minimization with Boltzmann sampling are given in the *Appendix* 12. This procedure was applied to 181 ligands carrying CM-1 charges [327] and to 16 ligands (out of the 181) carrying ESP-MNDO charges [331] so that also the influence of different charges-calculation methods on the MC-protocol could be evaluated.

3.2.5.2 Solvation docking

To investigate the effect of solvation on the minima search, 12 minimized complexes (result of the general docking procedure, *Chapter* 3.2.5.1) featuring CM-1 [327] charges (for details on the ligand structures refers to the *Appendix* 9 and 13) were solvated and further minimized. Water molecules including single hydrogen-bonded waters were therefore generated within a distance of 8 Å around the ligands using the standard water-cavity scan implemented in *Yeti* [248,352]. After a first minimization step, where only the water molecules were minimized, a full minimization of the complexes followed. The different parameters needed for a correct simulation, were set at the same values described in the docking general procedure.

For ligand **82** a different protocol to further study the effect of solvation was choose. In this case, the complex was solvated and minimized before applying the MC-protocol.

3.2.6 Molecular-dynamics simulations of protein-ligand complexes

The starting structure taken for MD-studies was always the most energetically favorable solution obtained from the docking protocol described in *Chapter* 3.2.5.1.

The pdb_ext file of the complex was then manually converted to the AMBER_pdb format. The ligand was extracted from this file and submitted to the ligand routine described in *Chapter* 3.2.3.2 with the difference that prior to solvation the ligand structure was re-integrated in the file of the E-selectin complex. The whole complex was then placed in a 14 Å wide pre-equilibrated TIP3P water box and counter-ions were added to the system. The topology- and the coordinates-file of the solvated complex were saved and used as an input for a first minimization step (cf. *Appendix* 14 minresCPL.in), where the solute was kept rigid. After minimization the complex was submitted to the same protocol described in *Chapter* 3.2.3.2 (*Figure* 32) to, first, optimize the position of the water molecules, and then relax the full system. Finally, instead of being submitted to the simulated annealing protocol as it happens for the ligand alone (cf. *Chapter*

3.2.3.2), the complex was heated up from 100 K to 300 K, where a sampling period of at least one nanosecond followed (cf. *Appendix* 15 md.in).

3.2.7 QSAR studies

3.2.7.1. Quasar

Receptor generation with *Quasar* [270–272,351] was carried out with 181 ligands, which were divided into a training- and a test-set. Indeed, four different set of 181 ligands were independently used for the derivation of a QSAR-model. Namely models were derived with: a first set of ligands presenting CM-1 charges [327] and coordinates obtained from a pharmacophore-based alignment of the ligands on sLe^x as found in the E-selectin crystal structure [166], a second one carrying the same charges but coordinates obtained after the MC-procedure, a third and a fourth one presenting ESP-charging [331] and either coordinates from the alignment- or MC-procedure.

After creating the three-dimensional surface, the ligands individual envelopes were generated using all of the six different induced-fit scenarios (module *Envelope*). In search of a consistent QSAR-model, which could predict the binding affinity towards E-selectin in good agreement with the experimental data, the following parameters were varied:

- Population size
- Solvation energy and partial-charge equalization
- E_{solv} attenuation factor
- Dynamic surface area
- H-bond radii
- Number of cross-overs

The obtained models were rated on the basis of q^2, p^2, and rms values. The best models were also subjected to a scramble test [338] to assess the sensitivity of the model towards the biological data.

3.2.7.2 Raptor

 For our best *Quasar* models (Q15 and Q21) *Raptor* [315,316,335] was used to generate others QSAR-model based on a different technology. In fact, the possibility of two different methods to similarly predict the affinities towards E-selectin of the same set of ligands, would bring us this consensus-scoring that is

nowadays emerging as state of the art in QSAR [273–282]. To achieve that, Raptor-models were generated using always ten individual models. For both ligands-sets the model generation was performed with and without enabling a threshold at 10 mM, to punish the prediction of weaker binding of the "non-binders" (IC_{50} > 10 mM) differently. Finally, the sensitivity towards the biological data was tested with a scramble test [338].

3.2.8 *De novo* design

The *de novo* design software *Allegrow* [318, 319] was used to develop new E-selectin antagonists based on the structure of the previous identified potent ligand **26** (*Figure* 35).
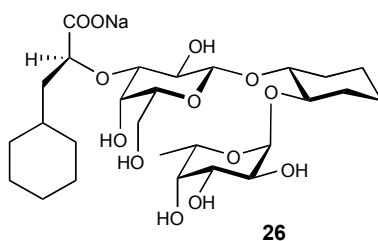
Figure 35: The chemical structure of compound **26**.

After a detailed analysis of the E-selectin binding site, it was decided to extend the ligand towards the pocket defined by the Lys111–113, Pro46, Tyr44, Tyr48, starting from one of the hydrogen atoms attached to the carbon atom at the position six of the galactose moiety of **26** (*Figure* 36).
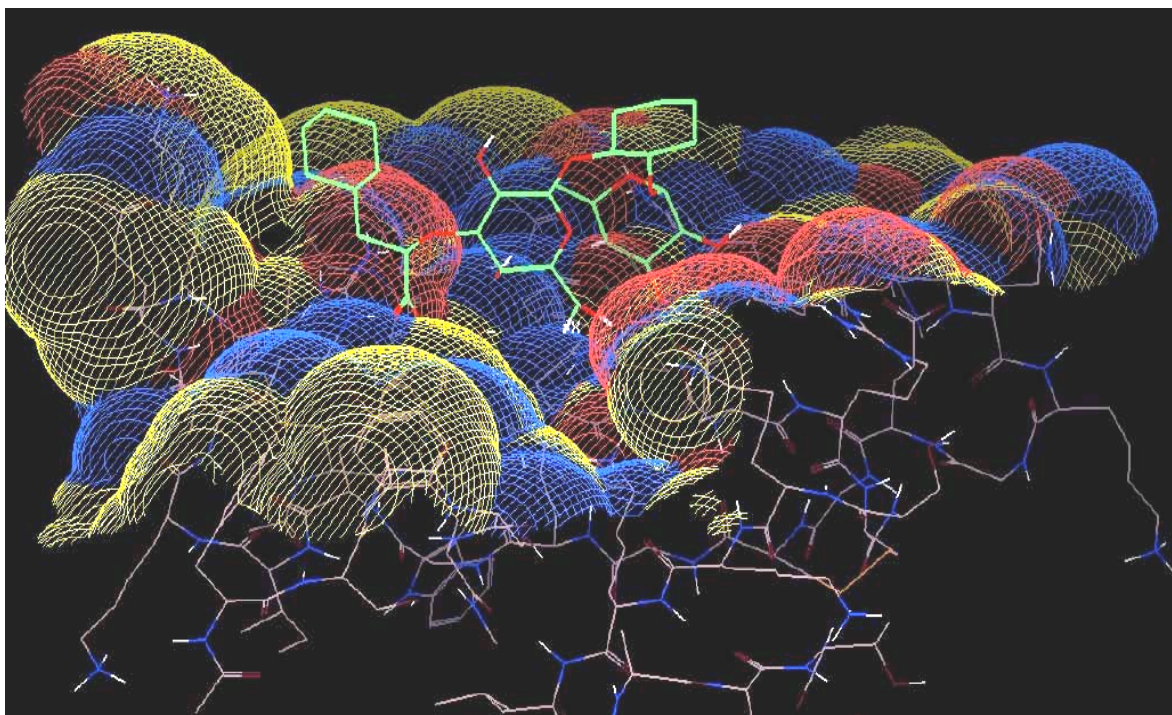
Figure 36: Compound **26** docked onto the surface of E-Selectin. One of the hydrogen atoms attached to the carbon atom at the position six of the galactose moiety was chosen as starting point for the generation of new ligands extending in the pocket defined by the amino acids Lys111–113, Pro46, Tyr44, and Tyr48.

The starting point for the design was structure **26** docked onto the E-selectin surface as obtained from the docking studies. This complex was read in *Allegrow* [318,319] and prepared for the "grow-procedure". In particular, all protein residues, excepted Lys111–113, Pro46, Tyr44, and Tyr48, which were colored blue (flexibility of the side chains), were colored red (rigid), whereas all the ligand atoms were colored green. The calcium ion was colored yellow and connected by so-called bonds of order zero, to its coordination partners (cf. *Chapter* 1.5.5). The atom-type of one of the hydrogen atoms attached to carbon at the position six of galactose was changed to X, thereby indicating to the software, which the "grow-atom" would have been. Finally, all residues located more than 12 Å from the ligand were deleted and the new structure saved. Subsequently, the grow-routine was started: 3 x 2,000 structures (using the three different fragment libraries at disposition: basic, normal, full) with a minimum of 6 atoms more added to a maximum of 12 atom added were generated and scored by the post-grow scoring-function [318,319]. The more potent ligands were visually inspected. A selection of the chemical more feasible compounds was done. In the end, a second "grow round" followed and the same analysis performed. Based on the

obtained structures the synthesis of a small library of compounds was planned (cf. *Chapter* 4.7.2.2).

3.2.9 Analysis of the results

The data produced was mostly analyzed with *Excel* and *Samoa* [337]. For the results generated with *AMBER* [322] (Conformational searches or MD of E-selectin complexes) a prior extraction of the data from the trajectory- and/or output-files with *Plotamber* [332] and *ptraj* [325] was necessary.

All the Φ/Ψ- and acid/core-plots were generated with *Samoa* [337], whereas the graphics tracking energy-, angles-, torsion, or distance-values over the time were obtained with *Excel*. Moreover, *Samoa* [337] permitted to calculate the ^3J-coupling constants and to extract important geometrical features (like torsion-angles values,…) from both (MC)JBW/SD- and MD-simulations. Finally, VMD [341,342] and Pymol [334] were used to produce the necessary movies and pictures for the analysis of the results.

4 Results and Discussion

4.1 Sampling conformational space of E-selectin antagonists: development of a new protocol

The conformational equilibrium state of a molecule in aqueous solution plays a determinant role for its affinity towards a target protein. It is assumed that the biological activity of a drug molecule depends on one single, unique conformation (the bioactive conformation), which is part of the low-energy conformational ensemble [217]. Moreover, it is known that compounds pre-organized in their bioactive conformation will show a higher affinity towards the target protein due the lower entropic cost they have to pay upon binding (cf. *Chapter 1.7.1*). The investigation and design of pre-organized E-selectin antagonists has become one of the major tasks at the Institute of Molecular Pharmacy. In 1997, Kolb and Ernst [161,168] presented a molecular-modeling concept for the investigation of the conformational behavior of E-selectin antagonists. This method, already described in *Chapters* 1.5.3 and 3.2.3.1, permits to semi-quantitatively determine the binding affinity of different compounds towards E-selectin by analyzing their conformational equilibrium in water: molecules showing a conformational distribution concentrated in the region of the known bioactive conformation of sLe^x, have been shown to be more active than less pre-organized ligands. For the following reasons, this method might not be perfectly suited for our system:

1) The fact that only 100 structures (out of several thousands as obtained from the conformational search) can be used as an input for the JBW-stochastic dynamics (SD) simulation represents a technical limitation. Consequently, the choice of the energy wells that will be investigated in the SD is biased (cf. Chapter 3.2.3.1, *Figure* 33). This restriction may lead an uncomplete sampling such as in the case of **71**. As can be taken from *Figure* 37(a), 600 conformations of **71** could be identified within 20 kJ/mol from the energy of the global minimum by the conformational-search protocol. From *Figure* 37(b) it becomes evident that, after the selection of the 100 energetically most favorable conformations, one of the local minima (one of the energy wells) has been removed; its environment will therefore not be sampled during the (JBW)-SD simulation, leading to incomplete or even incorrect results (compare *Figure* 37(c) with *Figure* 37(d), which is the results of a molecular-dynamics simulation with a protocol that shall be discussed below).
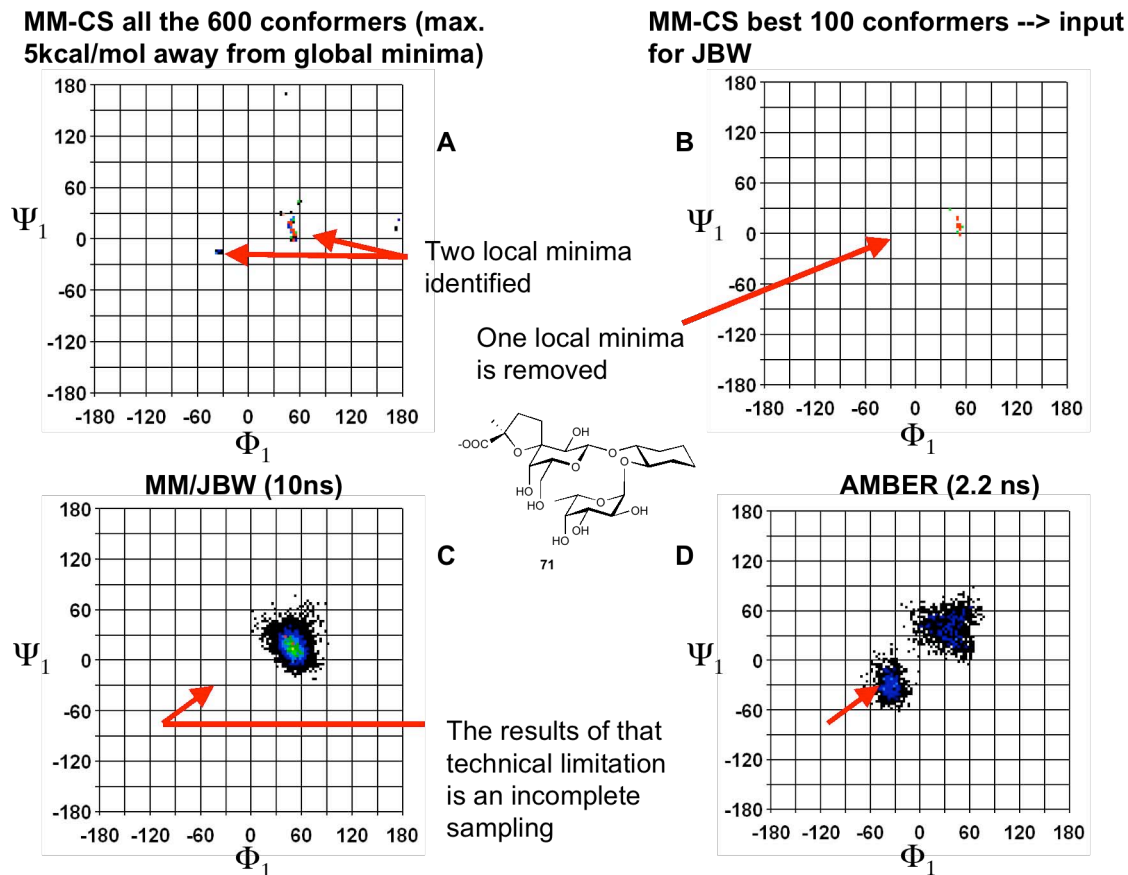
Figure 38: a) 600 conformations obtained from conformational search b) The 100 lowest-energy conformations chosen for the JBW-SD simulation c) Results of the JBW-SD simulation d) Results of the MD-simulation performed with *AMBER* [322].

2) An intrinsic solvation model (GBSA) [171] is used. This model is not particularly suited to handle charged molecules like E-selectin antagonists, which often are. Moreover, the directionality of the hydrogen bonds between the ligand and the solvent cannot be treated optimally, which often leads to the formation of intra-molecular hydrogen bonds that, in solution, are probably less important because of the possibility of interacting with the bulk water.

3) Unfortunately, the development of this technique has been discontinued in favor of the implementation of modern molecular-dynamics protocols.

We therefore developed an alternative concept that could integrate the MC(JBW)/SD approach in the conformational analysis for the design of new E-selectin antagonists. To tackle the aforementioned problems, we based the new procedure on the molecular-dynamics approach devised by Woods [343]. A detailed description of our new protocol is given in *Chapter* 3.2.3.2.

The main differences to the MC(JBW)/SD approach as established by Kolb and Ernst [161,168], are the explicit treatment of the solvent, the use of the GLYCAM 2000 parameter-set [344] specifically designed for molecular-dynamics simulations of carbohydrates within the *AMBER* suite [322], and the overall tailoring of the protocol based on molecular dynamics simulations instead of on a MC(JBW)/SD approach. The advantages of this technique are thought to be the ability to overcome energy barriers by applying a simulated annealing protocol [232], the better treatment of the solvation effects and of the electrostatic (the addition of counter ions balancing the total charge of the system is possible), and the use of a well established parameter set under continuous extension.

Some problems may still arise. In particular, due to the high number of rotatable bonds that are usually present in E-selectin antagonists, the completeness of the search cannot be assured in a reasonable amount of time.

To challenge the capability of the new protocol to reproduce results achieved with the MC(JBW)/SD approach and confirmed by the biological testing [161,168], the same set of ligands (**3**, **5**, **6**, **7**, **72**, **73**), that was used by Kolb and Ernst [161,168] to validate their molecular-modeling tool was submitted to the new protocol and the results compared (*Figure 38*). An excellent agreement of the results could be achieved with the exception of the simulation involving sLe^x. This very result will be discussed apart (cf. *Chapter* 4.2). Hence, it was decided to implement this new technique into our research as a standard method for the investigation of the conformational behavior of E-selectin in water.
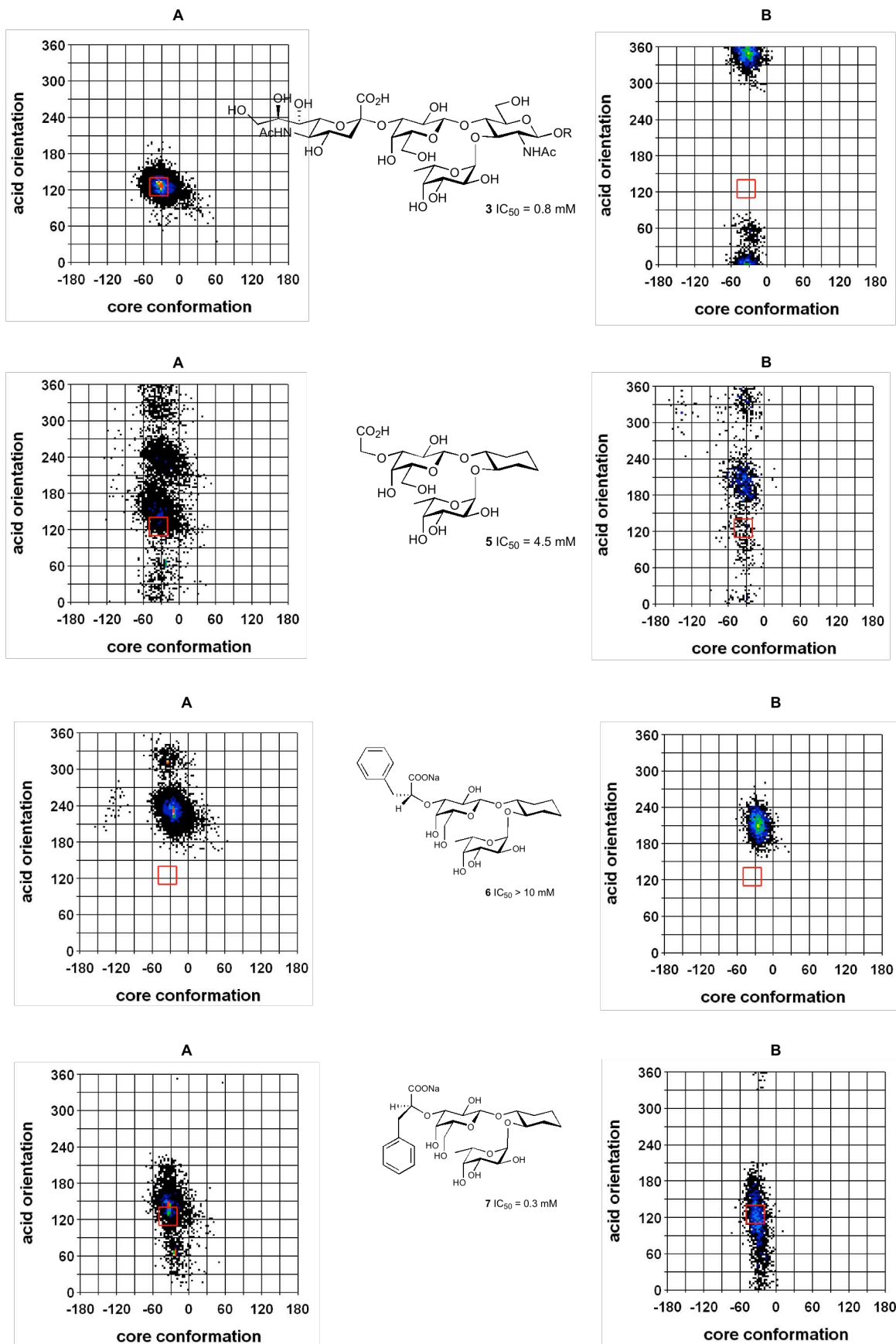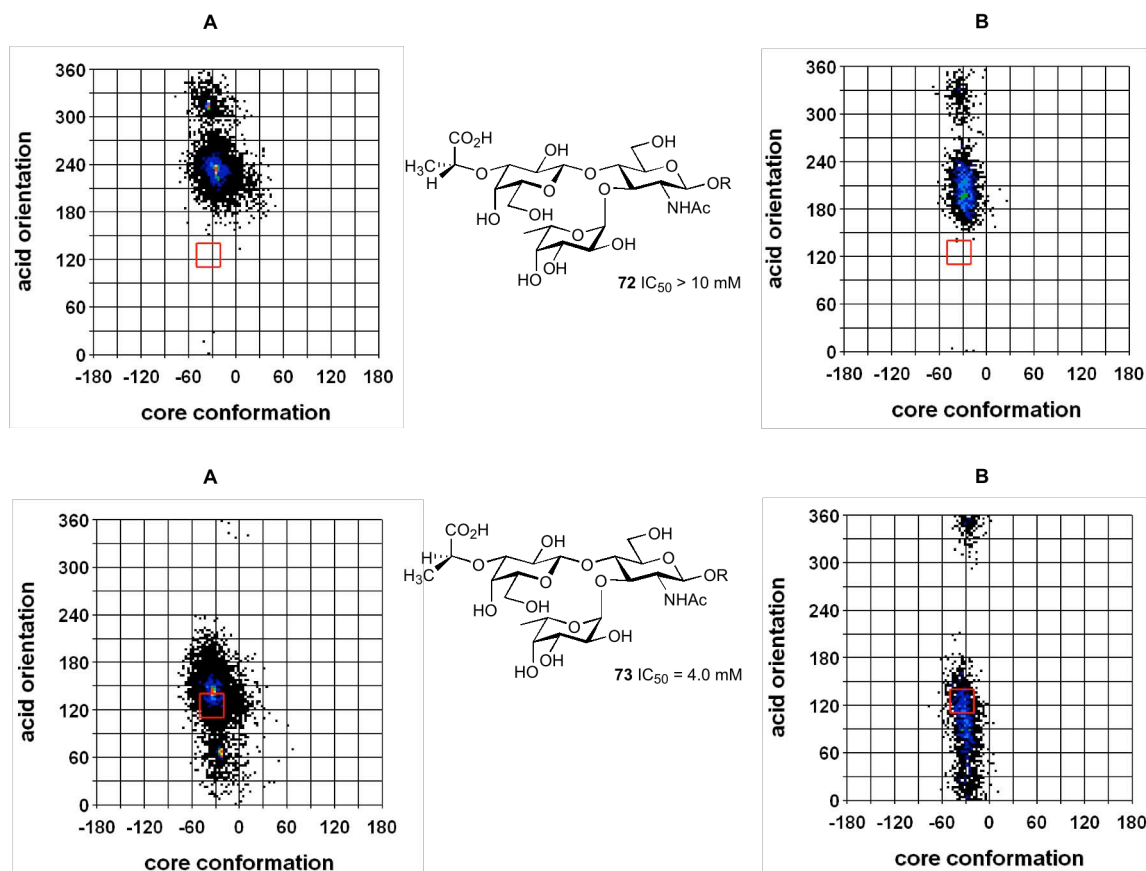
Figure 38: Comparison of the results obtained for the simulated conformational preference of the ligands **3**, **5**, **6**, **7**, **72**, and **73** by applying the MC(JBW)/SD- (**A**) or the AMBER-based-protocol (**B**).

Prior to that, the effects of different parameters on the simulation protocol were investigated. In particular, the coupling between the system and the heat bath (*taupt*-parameter) was of interest because it was known to directly affect the simulation results. Typically, values for *taupt* should be in the range of 0.5–5.0 ps, with a smaller value providing tighter coupling to the heat bath, therefore resulting in a less natural trajectory. Two different values of 1.0 and 4.0 were tested for both the heating- and the cooling-phase (cf. *Chapter* 3.2.3.2) on several ligands (**3**, **4**, **7**, **26**, **74**). The results obtained for **74** (*Figure* 39) are presented in *Figure* 40. By comparing the results of the four MD-simulations with the acid-core plot obtained by the MC(JBW)/SD simulation, it becomes evident that the default value of 1.0 produces the more consistent results. When instead a value of 4.0 is applied during the heating phase, the geometry of the cyclohexane ring is unrealistically distorted. The default value of 1.0, was therefore chosen throughout all of our simulations.
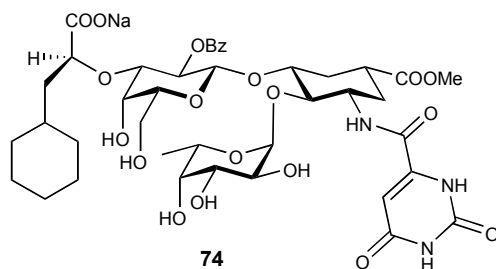
Figure 39: Chemical structure of compound **74**
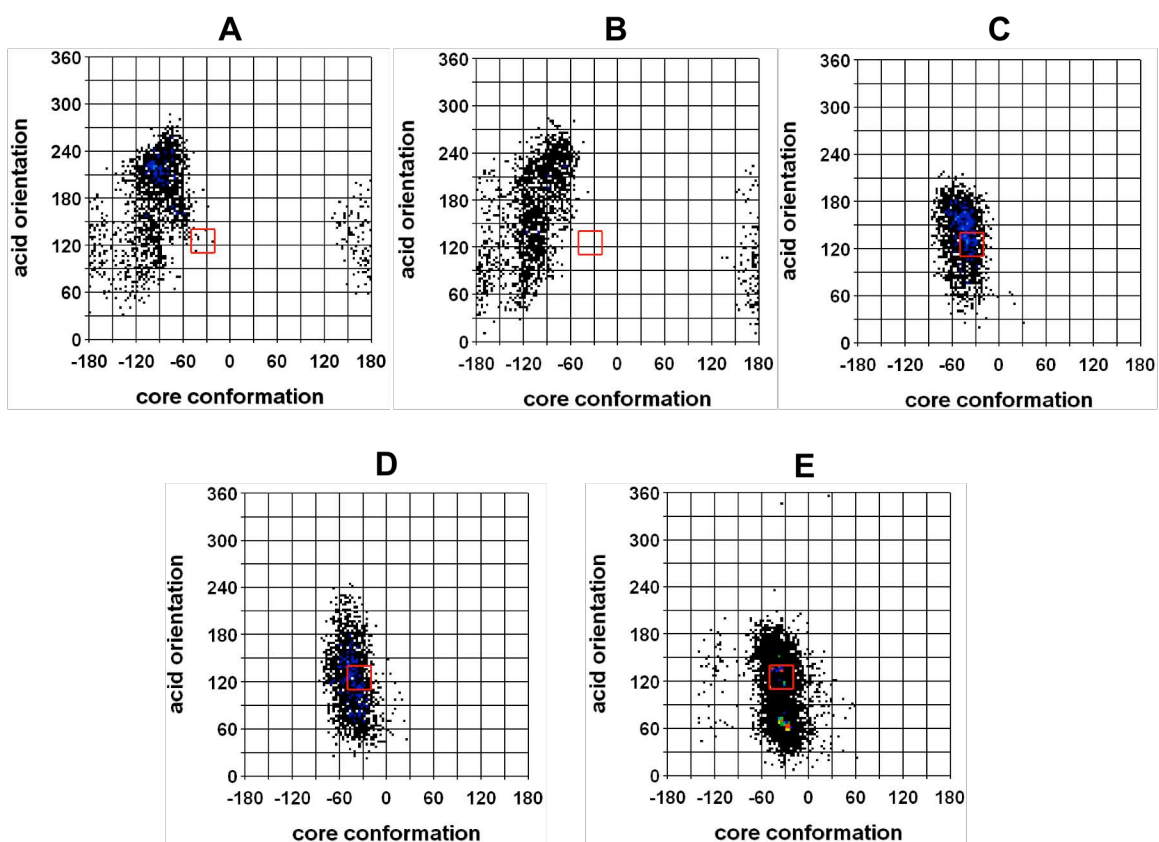
Figure 40: a) Simulation performed with the *taupt*-parameter set to 4.0 during the heating phase and to 1.0 during the cooling phase; b) Simulation performed with the *taupt*-parameter set to 4.0 in both phases; c) Simulation performed with the *taupt*-parameter set to 1.0 during the heating phase and to 4.0 during the cooling phase; d) Simulation performed with the *taupt*-parameter set to 1.0 in both phases; e) Simulation performed with MC(JBW)/SD-approach.

4.2 Investigation of the conformational space accessible by sLe^x

In the last two decades the investigation of the conformational properties of sLe^x (**3**) drew great interest [149–168]. Unfortunately, no consensual agreement has been found to date on the conformation of sLe^x in aqueous solution. As already introduced in *Chapter* 1.5.2.2, the dispute mainly focuses on the orientation of the sialic-acid moiety with respect to the Le^x-core substructure. While other opinions exist [153], most scientists agree on the fact that the Le^x-core substructure can be described by the following Φ–, Ψ–values ($\Phi_1 = 30° \pm 10°$, $\Psi_1 = 30° \pm 10°$; $\Phi_2 = 39° \pm 10°$, $\Psi_2 = 15° \pm 10°$) (cf. *Figure* 8) and that it doesn't alter its conformation upon binding. By analyzing more recent literature data [149–168], three main suggestions on the conformation sLe^x adopt in water are found. They all share a similar adjustment of Le^x-core but differ on the orientation of the sialic-acid residue:

1) $\{\Phi_3 = -180° \pm 10°, \Psi_3 = 0° \pm 10°\}$

2) $\{\Phi_3 = -70° \pm 10°, \Psi_3 = 0° \pm 10°\}$

3) $\{\Phi_3 = -100° \pm 10°, \Psi_3 = -45° \pm 10°\}$

Having two theoretical tools at disposition that produce consistent and correct results when investigating the conformational behavior of various E-selectin antagonists (cf. *Chapter* 4.1), we decided to perform the very investigation for sLe^x. As already discussed in the previous chapter, the MC(JBW)/SD- and the *AMBER*-based protocol, yielded different results (*Figure* 41).
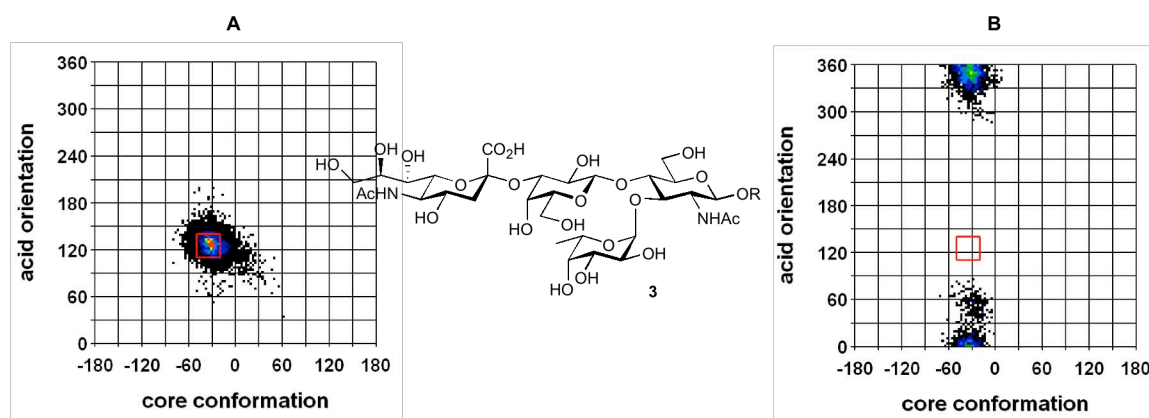
Figure 41: Simulated conformational distribution of sLe^x in aqueous solution as predicted by the MC(JBW)SD-protocol (**A**) and by *AMBER*-based simulation (**B**).

By analyzing them in more detail (*Figure* 42), it can be observed that the mismatch relies on a different conformational preference of the two methods for the orientation of the sialic-acid moiety. The *AMBER*-based simulation converges to the minima {$\Phi_3 = -180° \pm 10°$, $\Psi_3 = 0° \pm 10°$}, whereas the MC(JBW)/SD-protocol tends towards the (-)-gauche minima ({$\Phi_3 = -70° \pm 10°$, $\Psi_3 = 0° \pm 10°$}).
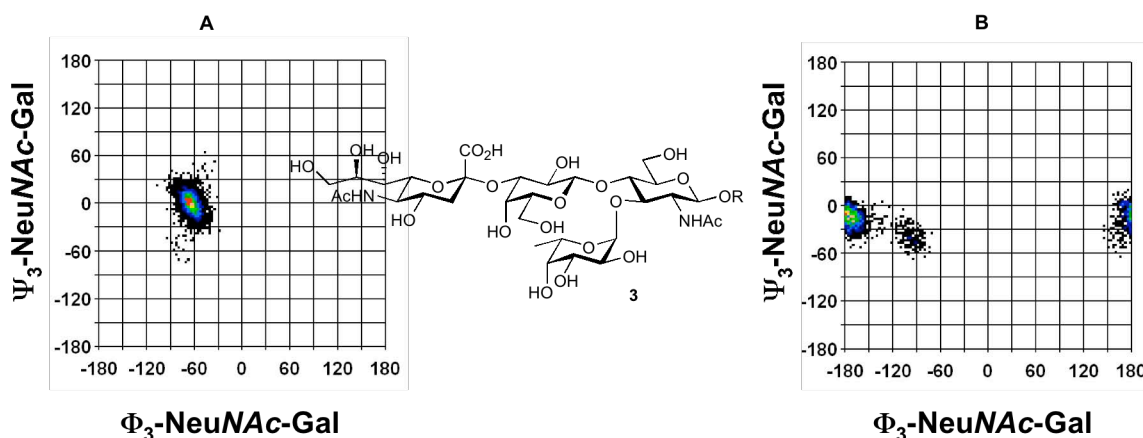
Figure 42: Simulated conformational distribution around the Φ_3– and Ψ_3-torsion of sLe^x in aqueous solution as predicted by the MC(JBW)SD-protocol (**A**) and by *AMBER*-based simulation (**B**).

Further investigation revealed that the potential around the torsion (C-EC-OG-CT, *Figure* 43), defining the orientation of the carboxyl acid thought to be essential for the binding was different.
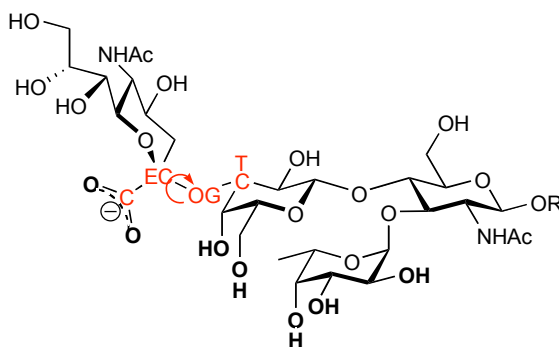
Figure 43: The torsion C-EC-OG-CT (GLYCAM 2000 [344] nomenclature).

As can be seen in *Figure* 44, both protocols yield a similar overall profile of this torsional potential but the energy barriers between the three minima ((+)-gauche, trans, and (–)-gauche) are quite different. Both methods identify in the (+)-gauche the most favorable orientation for the (C-EC-OG-CT) torsion to be assumed. However this conformation is never reached in the case of sLe^x due to steric hindrance (*Figure* 45a). When analyzing the second and third most favorable orientation (trans and (–)-gauche respectively), the trans state of the sialic acid is clearly favored over the (–)-gauche in the GLYCAM [344] parameter file by a factor of ten (ΔE = 1.4 kcal/mol).
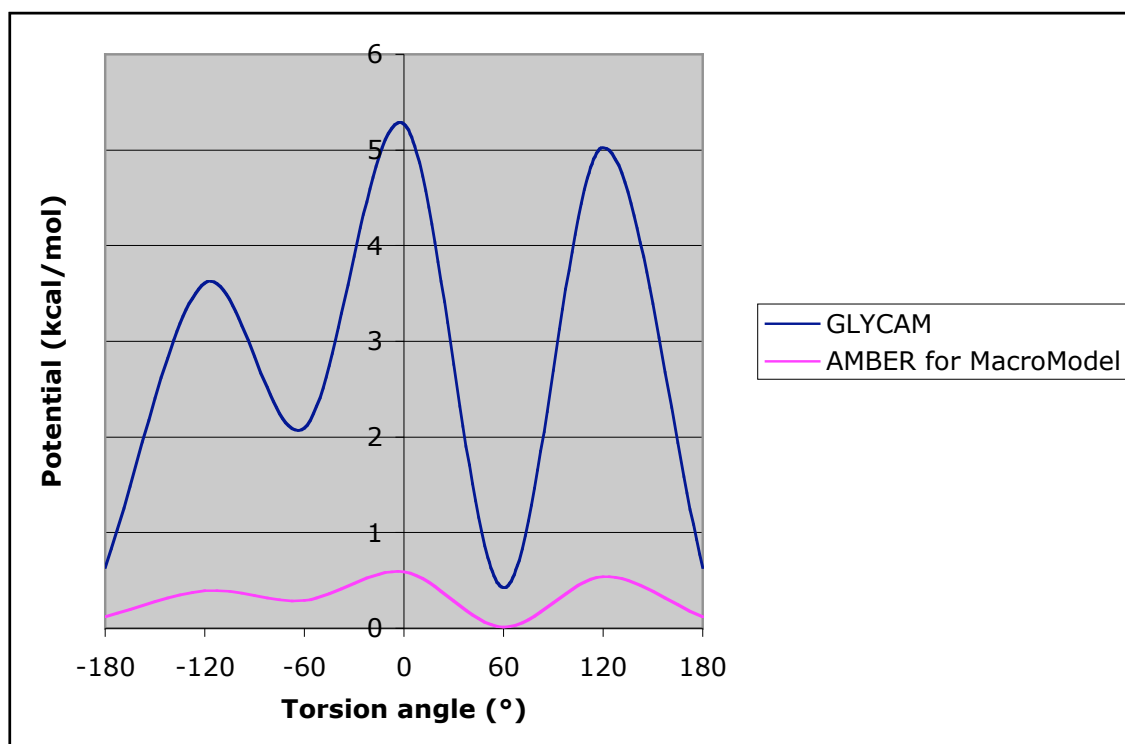
Figure 44: Energetic profile of the potential around the torsion (C-EC-OG-CT).

Even though this torsional potential is only one of the contribution to the total energy influencing the conformational behavior of the sialic-acid moiety, overweighting of the trans conformation would seem to be sufficient to explain the convergence of *AMBER*-simulations toward the trans minima. In contrast her to, in the *AMBER* force field [161,168,170] as implemented in *MacroModel* [169], the potential for the very torsion is much flatter, hence less favoring a particular minima. The global minima conformation ((–)-gauche conformation) as obtained with the MC(JBW)/SD-protocol (*Figure* 45b), is stabilized by an intramolecular hydrogen bond between the Gal–2OH and the Neu–7OH. This stabilization is not possible when the sialic acid assumes a trans orientation (*Figure* 45c) resulting in

a 2.0 kcal/mol higher energy. However, this interaction could also be an artefact generated by an overestimation of the importance of intramolecular hydrogen bonds when using an implicit-solvent approach. This would seem to be mitigated by the presence of real water molecules in explicit solvent models. Unfortunately, the different minima solutions obtained with our approaches reflect the two concepts (cf. above) and both are valid — albeit for different reasons. In fact, has already explained in *Chapter* 1.5.2.1, few and controversial experimental data [149-154] is available; a correct parameterization of the Sia-Gal torsions is therefore difficult.
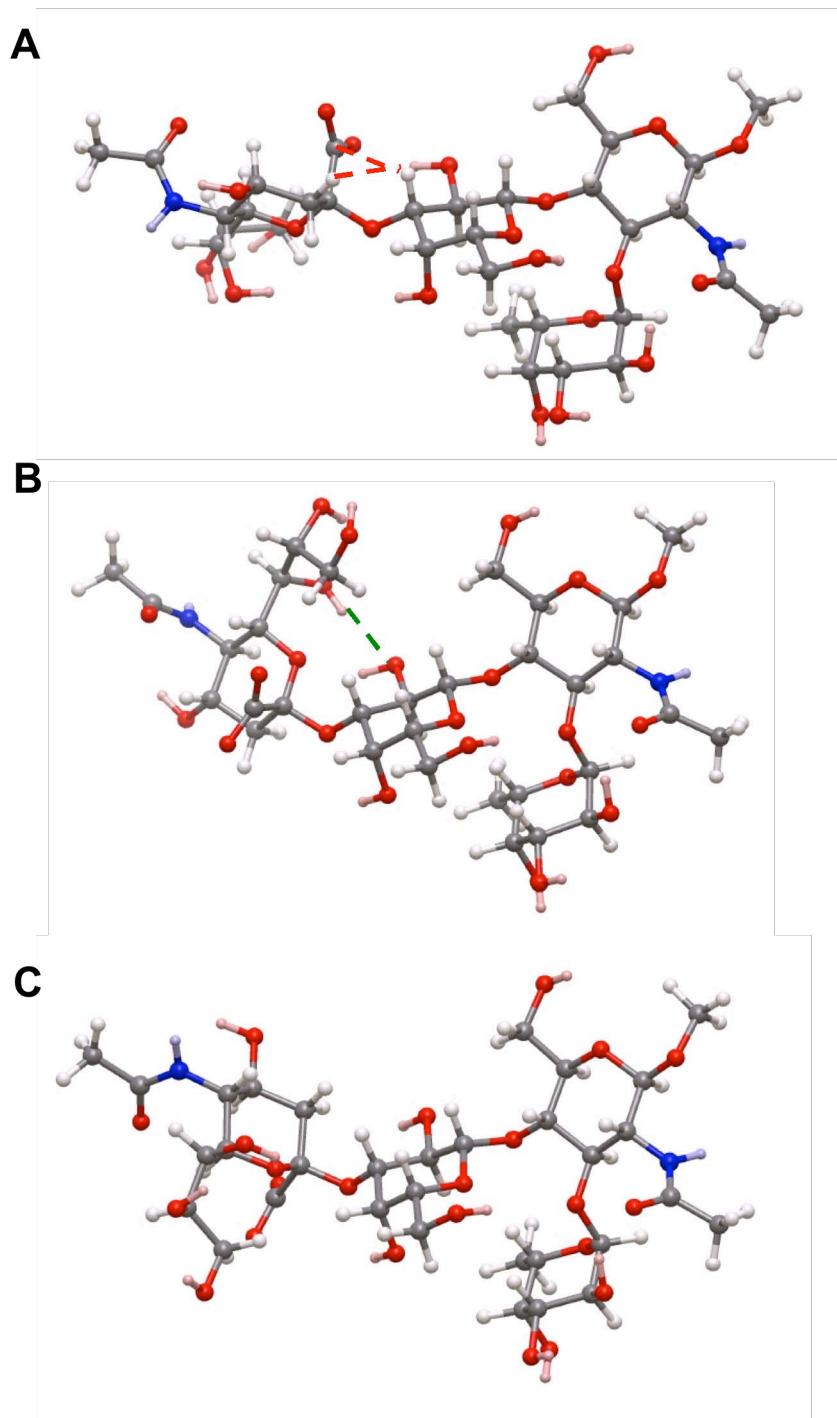
Figure 45: The three conformations of sLe^x (**3**) representing the different minima of the potential around the torsion (C-EC-OG-CT). a) (+)-gauche (red = the steric clashes); b) (–)-gauche (green = hydrogen bond of the Gal–2OH and the Neu–7OH); c) the trans orientation

In 2001, Bose et al. [159] published an extension of the Karplus equation for the calculation of $^3J_{C-C}$ coupling constants (cf. *Eq.* 2) specifically parameterized for carbohydrates. We applied this equation as well as the one previously derived by Tvaroska et al. [158] (cf. *Eq.* 1) for the calculation of the $^3J_{C-H}$ coupling constants of sLe^x from the two MD-ensembles and compared the calculated ^3J-values with the experimental data [167]. The results are given in *Table* 6.

Table 6: The ^3J-coupling constants calculated from the two MD-ensembles compared with the experimental data [167]. Green = good agreement, orange = poor agreement, red = disagreement. All values are given in Hz.

$^3J_{C-H}$	Experimental Data	sLe^x-JBW (gauche)	sLe^x-AMBER (trans)
FucH1-GlcNAcC3	2.8 ± 0.5	3.18	2.91
FucC1-GlcNAcH3	5.1 ± 0.5	5.03	4.91
GalH1-GlcNAcC4	2.8 ± 0.5	2.35	3.02
GalC1-GlcNAcH4	5.0 ± 0.5	5.4	4.9
SiaC2-GalH3	5.4 ± 0.5	5.51	5.4

$^3J_{C-C}$	Experimental Data	sLe^x-JBW (gauche)	sLe^x-AMBER (trans)
SiaC1-GalC3	1.9 ± 0.5	3.56	0.29
SiaC2-GalC2	1.8 ± 0.5	0.75	1.63
SiaC2-GalC4	<1.0	1.43	0.43
GalC2-GlcNAcC4	2.9 ± 0.5	3.55	3.55
GalC1-GlcNAcC3	2.3 ± 0.5	1.63	1.99
GalC1-GlcNAcC5	<1.0	0.55	0.48
FucC2-GlcNAcC3	3.3 ± 0.5	3.24	3.5
FucC1-GlcNAcC2	<1.0	0.37	0.2
FucC1-GlcNAcC4	2.2 ± 0.5	2.1	2.01

Both methods yielded ^3J-coupling constants in good agreement with the experiment, suggesting the validity of both approaches. However, one exception must be noted. Particularly with the $^3J_{C-C}$ coupling constant defining the orientation the (C-EC-OG-CT)-torsion, there is a disagreement. Interestingly,

both calculated values differ significantly from the experiment — an important finding indicating that in water sLe^x is neither represented by 100% trans conformers as suggested by the *AMBER*-based protocol nor by 100% (–)-gauche as proposed by the MC(JBW)/SD-procedure. Even when calculating the $^3J_{SiaC1\text{-}GalC3}$ for the conformation representing the energy minimum hypothesized to be the most frequently populated by Veluraja and Margulis [154] ({$\Phi_3 = -100° \pm 10°$, $\Psi_3 = -45° \pm 10°$}) a value of only 0.26 Hz is obtained. Again, this value differs from the experimental one (1.9 Hz $\pm$ 0.5 Hz). Hence, it may be concluded that it is probable that in water sLe^x exists as a mixture of different conformations and not as supposed to date [150–154, 161–163] by populating a single conformer. As a consequence, the idea to design mimics with a pre-organized structure, and in particular a pre-oriented acid moiety gained importance.

4.3 Comparison of two approaches of computing partial atomic charges: ESP-MNDO and CM-1[3]

For all ligands studied in this work, two methods, ESP-MNDO [331] and CM-1 [327] were applied to compute atomic partial charges. The idea behind was to analyze the impact of a different atomic charge model:

1. on the very binding mode to the target and

2. on the development of a QSAR-model for affinity estimation.

In general, the two methods yielded quite similar atomic partial charges. However some distinct differences were observed for certain atom types. As an example, *Figures* 46 show three representative ligands (**75**–**77**), where inequalities for the partial charge carried by some atoms is observed. Essential differences were found for nitrogen atoms (*AMBER* atom-type: N), chlorine atoms, and for the atoms within –C=N-fragments. The charge of chlorine is negative when applying the CM-1 [327] methodology and positive when using the ESP-MNDO [331] one. Moreover, according to CM-1 theory [327], the nitrogen atom in the C–N bond are negatively charged, whereas the carbon atom is positively charged. When ESP-MNDO-charges [331] are computed instead, both the nitrogen and carbon atoms of the C–N bond are slightly negatively charged. Thus compared to ESP-MNDO [331], CM-1 [327] calculations led to a more polarized N–C bond. The dissimilarities observed probably arise from the different philosophies behind the two approaches. Hence, ESP MNDO [329,331] tries to reproduce the electrostatic potential whereas CM-1 [326,327] the dipole moment.

---

[3] Results of this chapter were obtained as a part of a diploma work (M. Schmid, 2004) under my supervision.
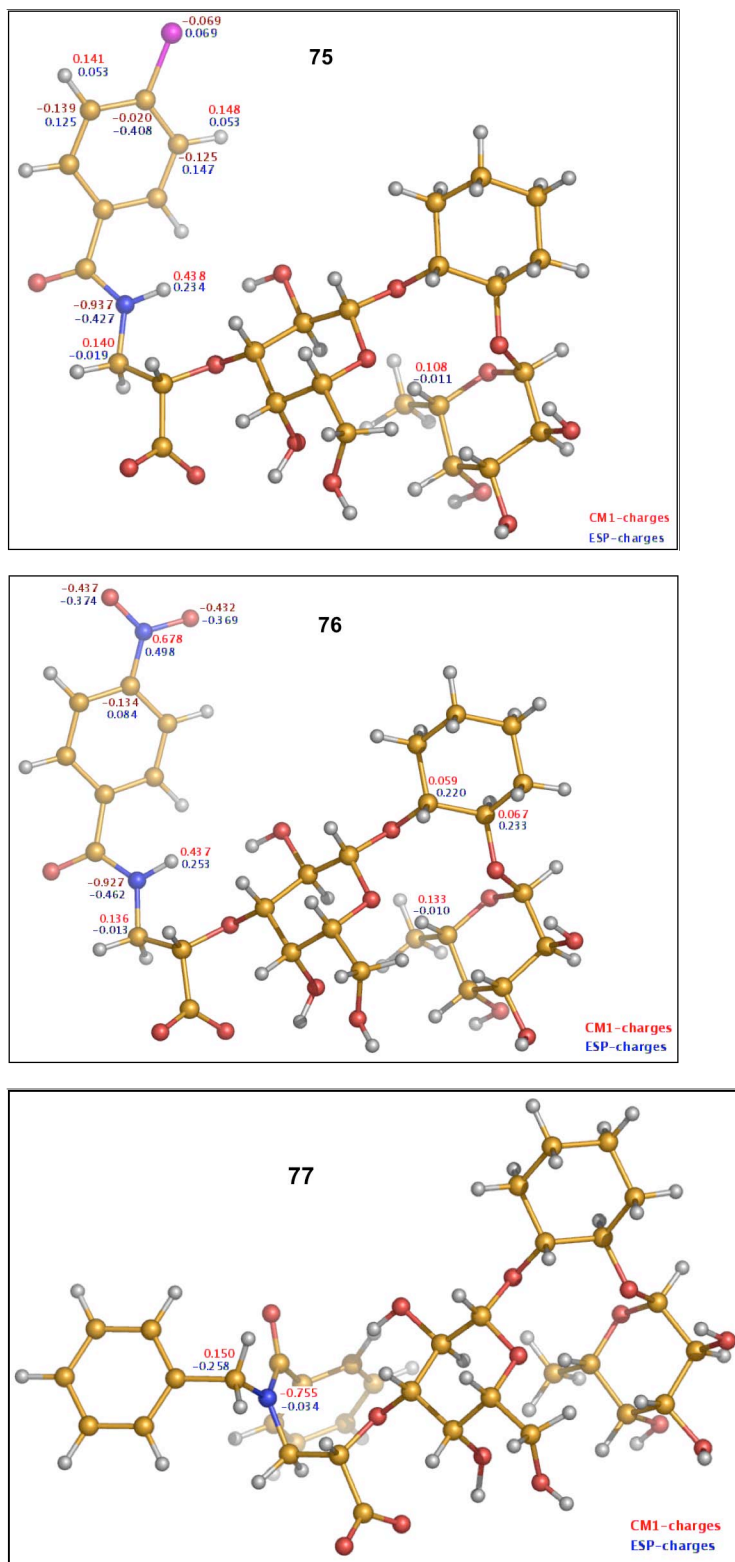
Figure 46: The observed differences in partial atomic charges are highlighted for ligands **75–77**.

4.4 Docking results[3]

Prior to start our docking simulations a careful evaluation of the available software for the planned task was performed. Different studies [265–268,348] have shown that each program has its strengths and its weaknesses. At the Institute of Molecular Pharmacy, different pieces of software (*AutoDOCK* [258], *QXP* [329], and *Yeti* [248,352][4]) were at disposition. We selected *Yeti* [248,352] because it had been specifically designed for the modeling of metallo proteins. Moreover, as shown by Schulz-Gasch and Stahl [265], when dealing with a protein bearing a high polar surface (e.g. E-selectin), software using a hard scoring function is needed. *Yeti* [248, 352], whose force field energy expression contains a directional term for hydrogen bonds, would seem to accomplish this requirement in a perfect way. Further, in some preliminary studies performed (cf. *Chapter* 4.4.1) prior to the final docking experiments, *Yeti* [248,352] correctly reproduced the binding mode of sLe^x as observed in the crystal structure [166] and showed only little dependence on "external" factors such as atomic partial-charge models or solvation. Finally, the fact that *Yeti* [248,352] allows for the flexibility of the amino-acids side chains of the protein during the MC-searches was considered an advantage against other software.

4.4.1 Docking preliminary studies

4.4.1.1 Reproducing the docking mode sLe^x

The first test to evaluate the performance of *Yeti* [248,352] was to dock sLe^x in its binding site to check if the binding mode identified by the crystal structure [166] could be reproduced. The protein and the ligands were prepared as described in *Chapter* 3.2.4. The ligand, bearing CM-1 atomic partial charges [327] was manually docked onto the E-selectin surface and submitted to the MC-protocol designed above (cf. *Chapter* 3.2.5). The energetically best solution (the ligand-protein interaction energy was used as a criterium) obtained by this protocol, superimposed to the position of sLe^x as found in the crystal structure [166] is given in (*Figure* 47). As it can be observed only minimal differences are visible. An overall rms deviation of only 0.5 Å calculated.

---

[3] Results of this chapter were obtained as a part of a diploma work (M. Schmid, 2004) under my supervision.

[4] The use of *Yeti* [248,352], which was provided by courtesy of the *Biographics Laboratory 3R* [328], is kindly acknowledged.
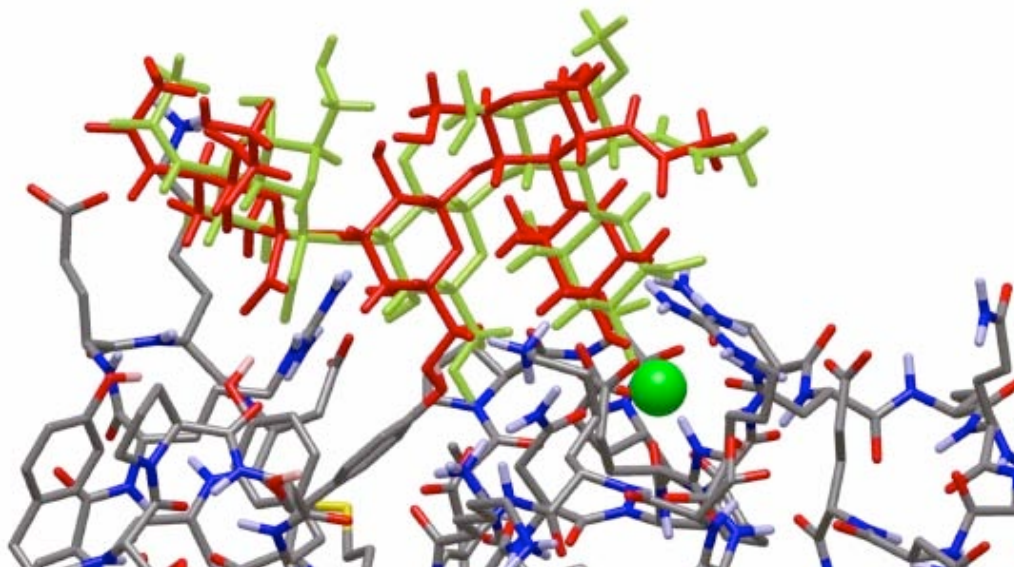
Figure 47: The docked structure (green) compared to the crystal structure (red).

4.4.1.2 Impact of the different partial-charge models on the docking results

To study possible effects of different partial-charge models on the docking mode, 16 ligands (cf. *Appendix* 16) bearing either CM-1 [327] or ESP-MNDO [331] charges were manually docked onto E-selectin and submitted to the docking protocol (cf. *Chapter* 3.2.5.1). The solutions are compared in *Figure* 48, which shows the superposition of the binding modes obtained with the two different partial-charge methods for ligands **78** and **79**. As it can be seen, the binding mode of the ligands seems not to be strongly affected by the use of different partial-charge methods. A possible explanation could rely on the fact that, in general, the partial atomic charge differences observed (cf. *Chapter* 4.3), involves atoms not being part of the pharmacophoric groups determining the binding mode. It was therefore decided to perform all further docking experiments only with ligands carrying CM-1 charges [327], which are calculated faster.
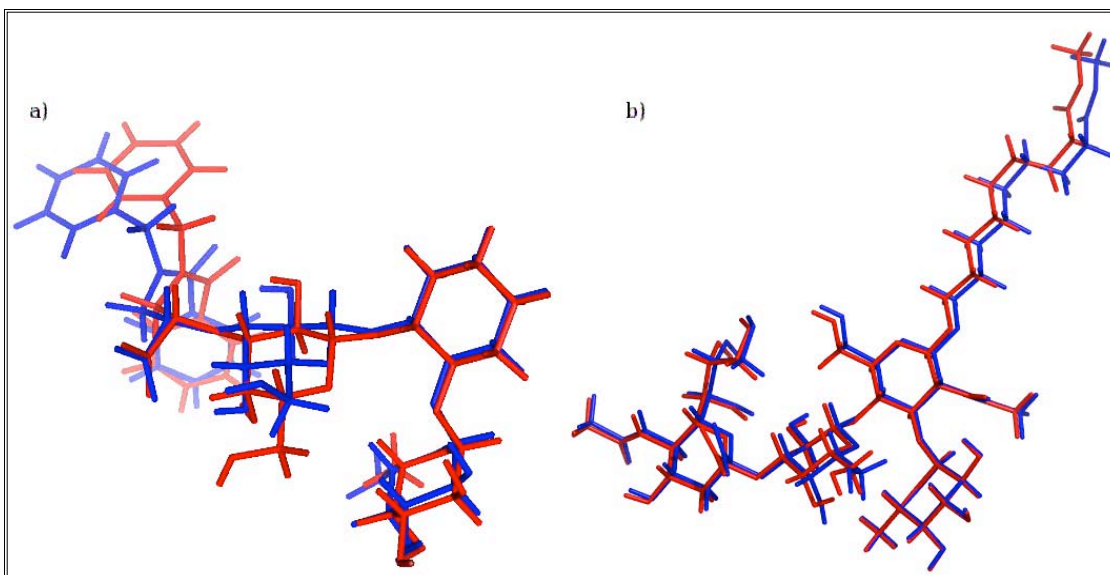
Figure 48: The docking mode obtained with the two partial-charge models: ESP-MNDO [331] (blue) and CM-1 [327] (red) for the ligands **78** (a) and **79** (b).

4.4.1.3 Impact of solvation on the docking results

To check the influence of solvation on global minimum obtained from the MC-protocol (cf. *Chapter* 3.2.5.1), 12 MC-minimized ligands bearing CM-1 charges [327] (cf. *Appendix* 13) were explicitly solvated as described in *Chapter* 3.2.5.2. A superposition of the minimized conformations before and after solvation provides information on the influence of the solvent on the global minimum identified. The compounds used in this study vary in lipophilic portions, backbone, as well as on their binding affinity to E-selectin and therefore represent a good structural subset to study the effect of solvation on all antagonists. In *Figure* 49, the superposition of the ligand conformations from **80** and **81** is depicted. The deviation between the ligand structures before and after the solvation is small (rms = 0.2 Å). A further investigation on the effect of the solvent on the minima search was carried out using **82**. This ligand was docked in the binding site but this time the MC-protocol was applied only after solvating the protein-ligand complex (cf. Chapter 3.2.5.2). As a result, the obtained conformation and the orientation of the ligand showed the flipping of the hydrogen bond of Gal–2OH and Gal–6OH and subtle deviations for the cyclohexane puckering and the carboxylate compared to the MC-minimized conformation without water (*Figure* 50). This result confirms our hypothesis that the influence of the solvent is minor and that the optimized protein-ligand interactions for this system seem to be superior to the effects of water molecules.
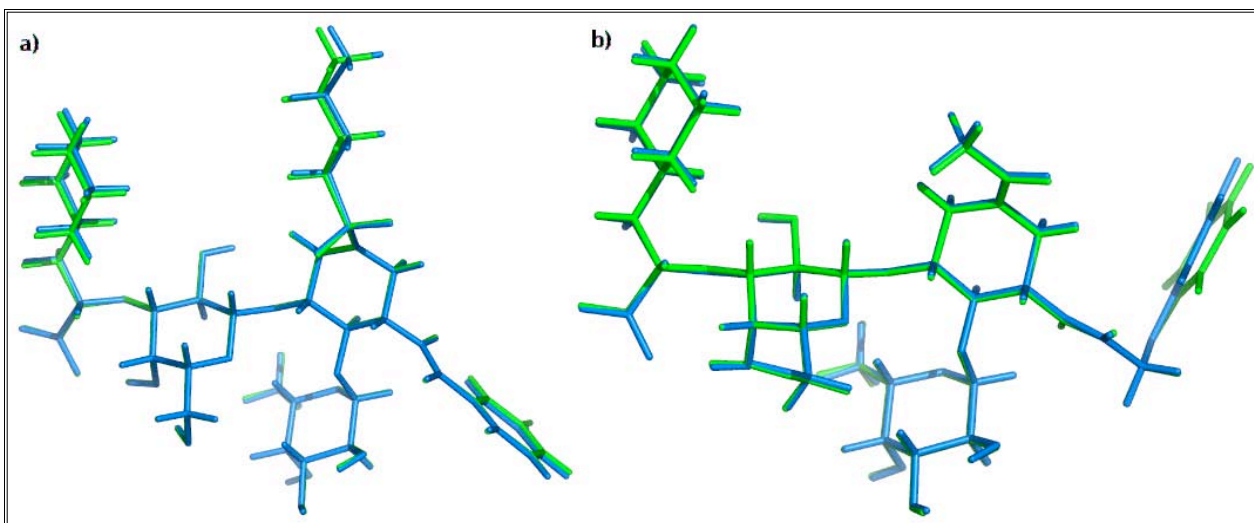
Figure 49: Superposition of the global-minimum solution obtained for the ligands **80** (a) and **81** (b) with (blue) and without (green) explicit solvation.
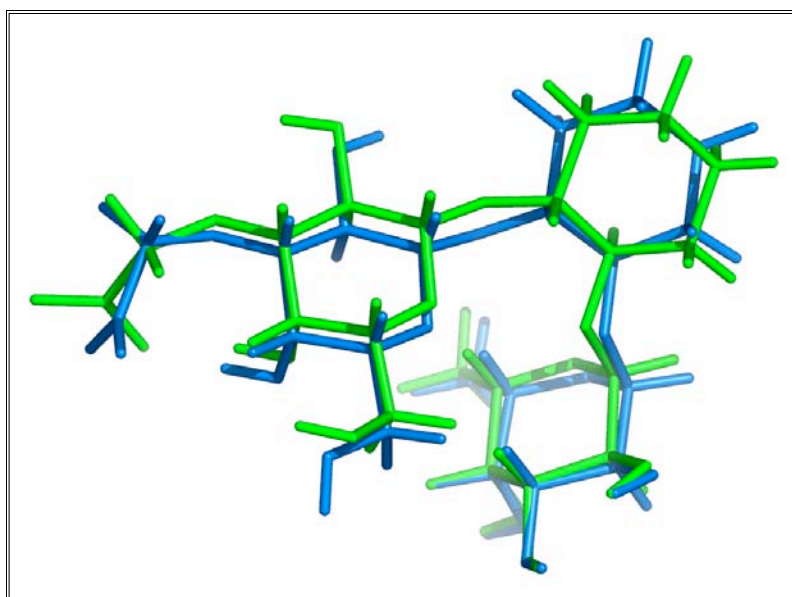
Figure 50: Superposition of the global minimum solution obtained for the ligands **82** with (blue) and without (green) explicit solvation.

4.4.2 General docking results

Our simulation yielded a binding pattern similar for all ligands, showing a binding mode similar to the one of sLe^x as found in the crystal structure [166]. This was expected since all compounds are quite similar to sLe^x (cf. *Appendix* 9). By looking at the superposition of the 200 conformations of each molecule populated during the sampling procedure (cf. *Chapter* 3.2.5.1) only slight differences could be observed (*Figure* 51). They frequently arise from the different orientation of side chains or from a slight translation (0.1-0.4 Å) of one of the sugar rings (*Figure* 52). Thus, it appears that E-selectin antagonists prefer a very conserved and unique binding mode. In particular, the importance of pharmacophoric groups (Fuc–4OH, Fuc–3OH, Gal–4OH, Gal6–OH and the acid moiety of Sia) could be confirmed. Hence, the absence of one those groups lead to lower E$_{lig\text{-}rec}$ (data not presented here).

As, the *Yeti* "scoring function"[5] [248,352] was not sensitive enough to distinguish between ligands having similar affinity towards the protein. It should be noted that most of the modifications leading to higher affinity towards the target protein (for example the introduction of hydrophobic substituents at the 2 position of the Glc$N$Ac moiety) seems rather to be oriented towards the solvent and should therefore only have little influence on the affinity on an enthalpic level. Instead, their role could be important on an entropic level by liberating a larger amount of protein-bound water molecules upon binding and by stabilizing the bioactive conformation of the ligand. Hence, for having access to a fast *in silico* estimation of the affinity towards E-selectin, the development of a QSAR-relation taking in account also solvation effects became mandatory (cf. *Chapter* 3.2.7).
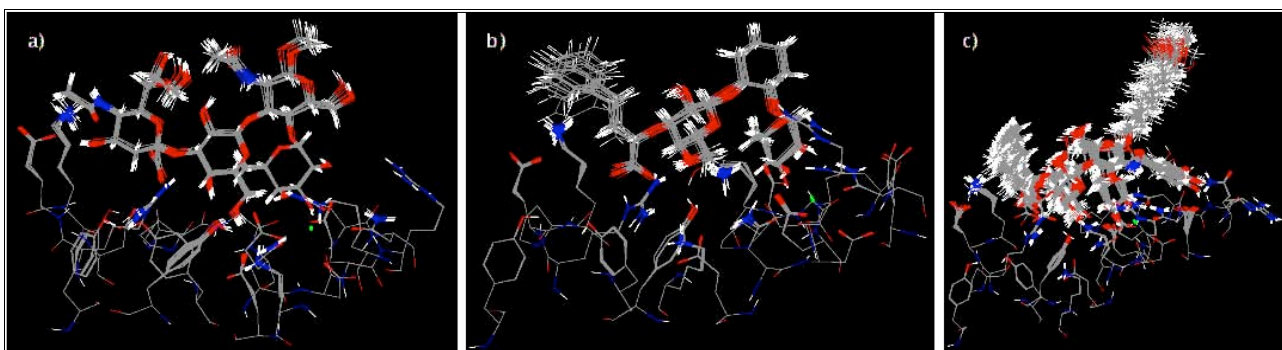
Figure 51: Graphical representation of the 200 sampled conformations (MC) for the ligands a) **83**, b) **84**, c) **85**.

---

[5] The weighting of the different energy components was changed to the values described in *Chapter* 3.2.5.1.
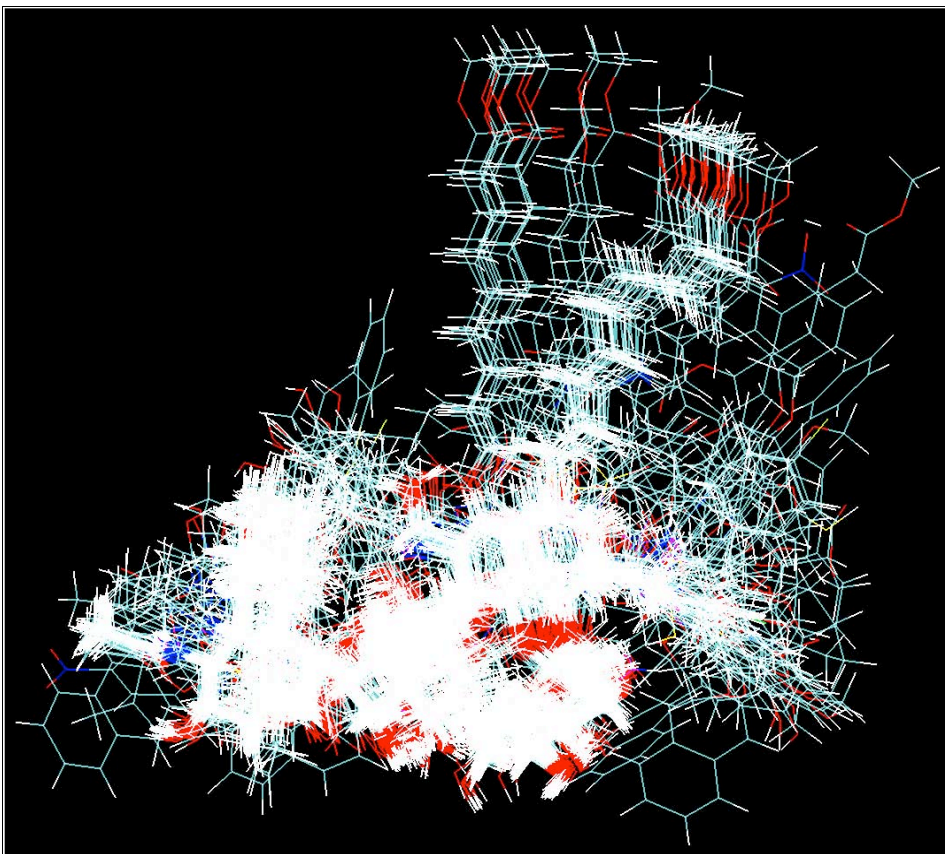
Figure 52: Superposition of the 181 ligands used for the docking- and QSAR-experiments after MC-minimization. Each ligand is represented by the lowest energy confomer.

4.4.3 Reverse docking

It is known from the literature [175,176] that the introduction of hydrophobic substituents at the 2 position of the Glc*N*Ac moiety clearly enhances the affinity of the ligands towards E-selectin (cf. *Chapter* 1.6.4), although reason for this remains unclear [176]. To date, two hypotheses have been put forward: the first, discussed in the *Chapters* 4.4.2 and 4.7.1 bases on entropic considerations, the second on an different docking mode. Both were first proposed within the framework of this work.

One possibility to explain an enhanced activity of a compound towards the target protein is that this particular molecule shows additional contacts with the protein, when compared to others ligands, thereby giving rise to stronger enthalpic contribution to the binding affinity. However, for compounds carrying hydrophobic substituents at the 2 position of the Glc*N*Ac moiety, no such evidence is detectable. Hence, the docking poses obtained out of the MC-protocol for those

compounds show the hydrophobic substituents rather orientated towards the solvent (*Figure* 53).
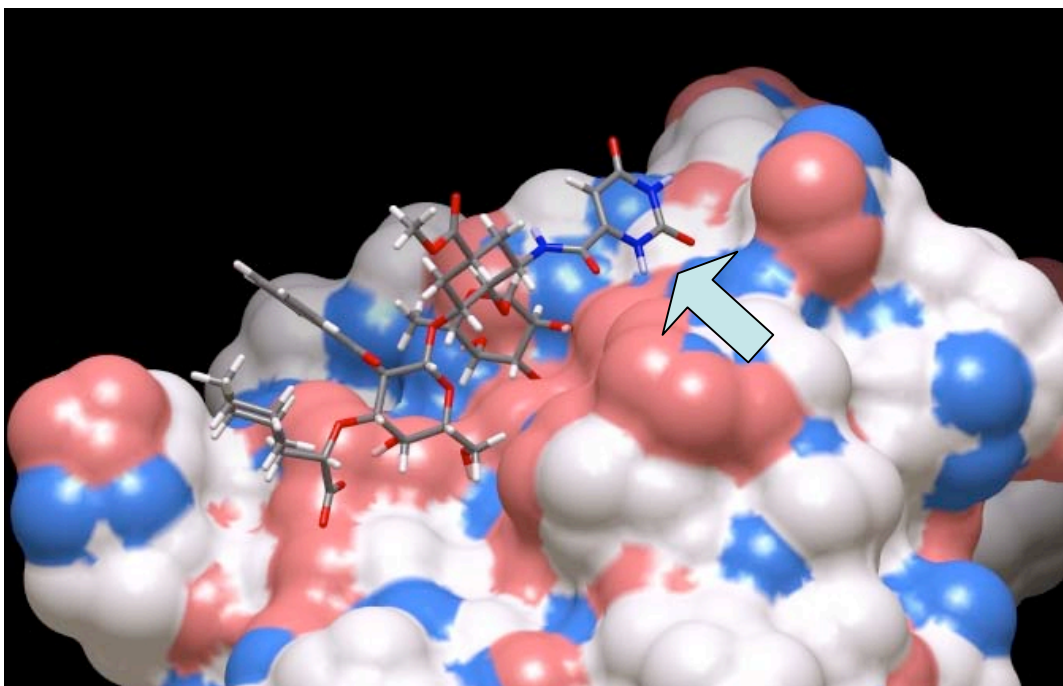
Figure 53: Compound **74** docked onto E-selectin. The hydrophobic substituent at the 2 position of the Glc*N*Ac moiety is orientated towards the solvent.

NMR experiments [349] indicated the presence of NOE-contacts between the hydrophobic moiety of ligand **68j** and the protein. Therefore, the hypothesis of a different docking mode for ligands carrying hydrophobic modification at the 2 position of the Glc*N*Ac moiety was considered.

First, the binding region of E-selectin was screened for regions, which could interact with the hydrophobic substituents attached at the 2 position of the Glc*N*Ac moiety. Three of those regions could be identified:

a) Tyr44-Pro46-Tyr48

b) Lys111-Lys112-Lys113

c) Ala77-Pro78-Met103

It was then attempted to dock the compounds bearing a hydrophobic moiety linked at the 2 position of the Glc*N*Ac residue with these rest being in close proximity of one of the identified regions on the E-selectin surface. Plausible binding modes, that could still explain the calcium-dependent binding of the E-selectin antagonists to the protein [81,86,92] and the results of the studies

investigating the structure-activity relationship of E-selectin antagonists [141–148], could only be achieved when positioning the hydrophobic moieties next to region a). The energetically and geometrically most favorable solution, obtained for ligand **74** (a complete list of the ligands studied can be found in the *Appendix* 17), is presented in *Figure* 54. As can be observed, in this different docking mode the molecules would bind to E-selectin, in an orientation that can practically be obtained by flipping the ligands of 180° about their longitudinal axis compared with the previously assumed binding mode. Therefore this new docking mode is referred to as "reverse docking mode".
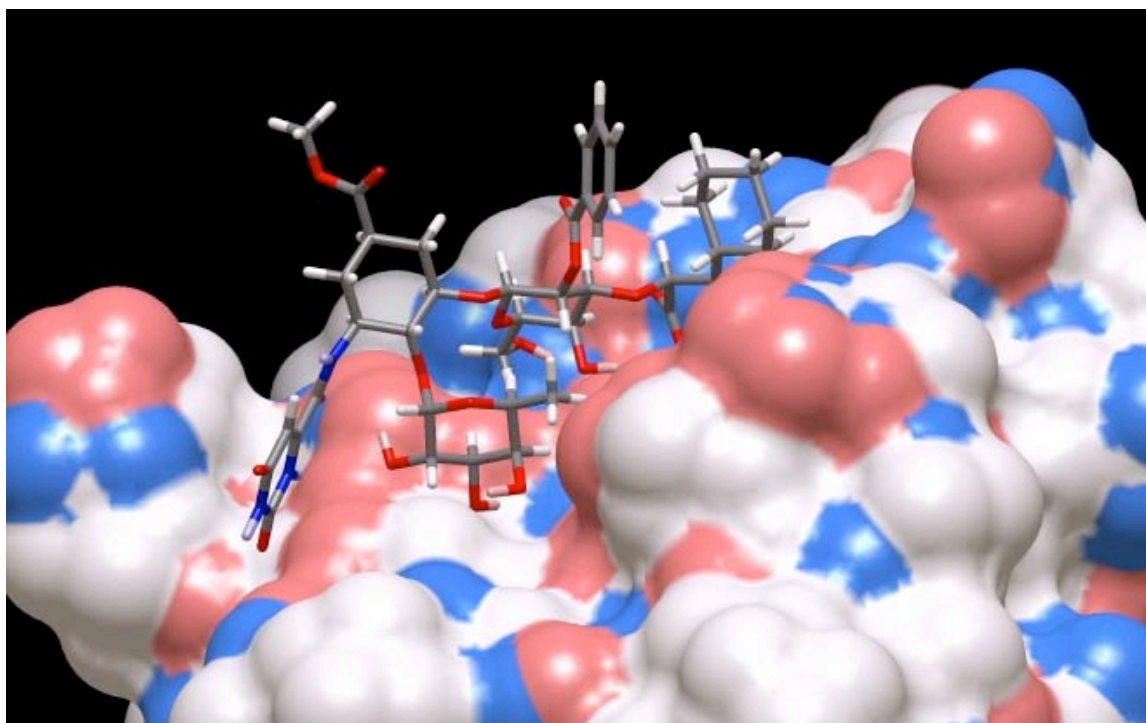
Figure 54: Compound 74 docked in the so-called "reverse docking mode".

In the reverse docking mode, the coordination to the calcium ion would be completed by the carboxyl group of the sialic acid residue (bidentate) of the ligand, where as the hydrogen bond to Glu92 would be guaranteed by the Fuc–3OH (as a donor). In addition, other hydrogen bonds between the Fuc–4OH (acceptor) and the lysines 111 and 113, as well as between the Gal–4OH (acceptor) and Asn105, and finally between the Gal–6OH (donor) and Glu80, are observed (cf. *Figure* 18). In conclusion, this docking mode fulfills all structural requirements and moreover seems to give a suitable explanation (additional hydrophobic contacts with the amino-acids Tyr44-Pro46-Tyr48) to the enhanced activity of compounds bearing hydrophobic substituents at the 2 position of the Glc*N*Ac moiety. It is, however, worth to note, that the "interaction" energy between the ligand and the protein ($E_{prot-lig}$, Yeti [248,352]) in this "reverse

binding mode" is smaller (2.0 to 10.0 kcal/mol) than when assuming a normal binding mode. Consequently, even though the "reverse binding mode" would seem to plausibly explain the role of the substituents at the position 2 of the GlcNAc moiety on both a structural and energetic level, the "normal docking mode" is still the more probable for this class of compounds. A definitive insight into the true binding mode of such ligands could only be reached through an X-ray structure of E-selectin co-crystallized with ligand **68a-j** or **74**. This project is part of an ongoing thesis at the Institute of Molecular Pharmacy.

4.5 Molecular-dynamics simulations of protein-ligand complexes

To further explore the interaction between E-selectin and its ligands, different molecular-dynamics simulations of protein-ligand complexes were performed. As starting point for the MD studies, the corresponding best solution from the MC-sampling was selected. The ligands for the MD studies were chosen to comprise a representative subset of all the 181 E-selectin antagonists previously docked to the target protein (cf. *Appendix* 18).

MD simulations allow for;

1) A more dynamic characterization (i.e. time dependent) of the protein-ligand interactions responsible for ligand binding. The idea was to gain more insight into the strength and lifetime of those interactions through the analysis of many MD trajectories. The gained information should have been implemented in the design strategy of new mimics.

2) The assumption that complexes of compounds having a higher affinity towards E-selectin would dissociate slower during the MD simulation than complexes formed by the protein and weak binders.

3) The possibility to identify the favorable docking mode for ligands bearing hydrophobic substituents at the 2 position of the GlcNAc moiety by analyzing the lifetime and interaction pattern of the two different docking modes.

The stability of the protein-ligand complexes did not correlate with the activities of the E-selectin antagonists but rather with their size. Larger ligands, independently from their affinity towards the target protein, displayed a longer residence time in the binding site when compared to smaller compounds. This is observed as well in the normal as in the reverse docking mode, making thereby difficult to identify one of the two docking mode as more probable. The longer "stability" of the complex formed between larger ligands and E-selectin might be explained with slower diffusion rate of heavier molecules.

With the MD simulations we gained some insight into the dynamical nature of the protein-ligand interactions. In particular, it could be observed that, as already postulated by Somers *et al.* [166], only few contacts, mainly of electrostatic

nature exist between the ligands and the protein. Moreover, it could be distinguished between long-lasting contacts and transient ones. To the latter category belong the hydrogen bonds formed by Gal–4OH and Gal–6OH with the side chains of Tyr44 and Glu92. Hence, this hydrogen bonds last for few picoseconds but are continuously broken and reformed. Instead, the interaction between the fucose moiety and the calcium ion as well as the salt bridge between the carboxylic group of the sialic acid moiety and Arg97, generally last for longer periods, and are always the last contacts to break upon the dissociation of the complex. These findings correlate well with the results of an atom-force microscopy experiment [350], which identified in the interactions of the sialic acid moiety and of the fucose moiety with the metallo-protein, the two main barriers in the dissociation process. This agreement supports the importance of these two interactions, giving further evidence in keeping the sialic acid and the fucose moieties or mimics thereof as central elements in the design of new E-selectin antagonists.

4.6 QSAR-models[3,6]

One of the more recent goals of the Institute of Molecular Pharmacy was to develop a QSAR model able to predict the binding affinity of newly designed E-selectin antagonists in a fast and reliable way. To date, no such model has been developed. The starting point for this project was given by the availability of an in-house database including some 300 chemical structures and their affinities towards E-selectin, and of two pieces of software (*Quasar* [270–272,351] and *Raptor* [315,316,335])[7] based on multidimensional QSAR.

4.6.1 Ligand selection

We decided that all ligands used for the QSAR studies had to satisfy some structural and charge criteria. In particular, it was chosen to include only ligands with an identical overall charge (e.g –1, as it is the case for sLe^x). Hence, in *Quasar* [270–272,351], the binding affinity is calculated as follow:

$E_{binding} = E_{(force field)} + E_{polarization} − E_{(ligand desolvation)} − T\Delta S_{binding} − \Delta E_{(internal strain)} − E_{(induced fit)}$

(**Eq. 4**)

---

[3] Results of this chapter were obtained as a part of a diploma work (M. Schmid, 2004) under my supervision.

[6] In this chapter the ligands will be discussed according to their three-letter code (cf. *Chapter* 3.2.4.2 and *Appendix)*

[7] The use of *Quasar* [270–272,351] and *Raptor* [315,316,335|, which was provided by courtesy of the *Biographics Laboratory 3R* [328], is kindly acknowledged.

and large deviation in the solvation energy of the individual molecules, as obtained for differently charged species, would have a considerable impact on the binding-site model. To avoid such an influence originating from different total charges, only structures with a total charge of -1 were taken in consideration.

Second, ligands, which could form tautomers, were excluded. These molecules are able to change their physicochemical properties — essential for their binding mode and for the modeling of the binding-site — depending on the tautomeric structure. To include this class of compounds supplementary docking experiments would have been required. Therefore, the QSAR studies were preformed with a smaller selection of ligands (e.g.181 structures).

4.6.2 Ligand-set selection

As the *best* partial-charge model cannot *per se* be identified, we decided to use both sets of ligands bearing either CM-1 [327] or ESP-MNDO as input for *Quasar* [270-272,351]. In addition, the so-called pharmacophore-based alignment and the receptor-based alignment [272,333] (cf. *Chapter* 4.6.3.2). In summary, the search for a predictive QSAR model explaining the structure-activity relationship between E-selectin and its ligands, was started with 4 sets of ligands (*Table* 7).

Table 7: Definition of the four sets used for the derivation of a QSAR model with *Quasar* [270–272,351].

Set number	Set name	Partial charge method	Alignment
1	CM-1 align	CM-1	Pharmacophore-based
2	ESP-MNDO align	ESP-MNDO	Pharmacophore-based
3	CM-1 MC	CM-1	Receptor-mediated
4	ESP-MNDO MC	ESP-MNDO	Receptor-mediated

As described, preliminary studies showed that docking was not influenced by which charge calculation method was applied. Set 4 was therefore obtained out of set 3 by exchanging CM-1 charges [327] against ESP-MNDO [331].

4.6.3 Alignment strategies

4.6.3.1 Pharmacophore-based alignment

The least-square technique for superpositioning of corresponding atom positions is the most widely used methods [302]. Two molecules are superimposed by minimizing the rms distances between the corresponding atom pairs in the molecules. The underlying idea is that by superpositioning three or more atom pairs an alignement of the pharmacophoric groups would be obtained. In our case, the ligands were superimposed onto the conformation of sLe^x as observed in the crystal structure of the sLe^x/E-selectin complex [166] using the three glycosidic oxygen atoms of sLe^x as reference atoms. The alignment was then further syncronized by reorienting some hydroxyl groups (cf. *Chapter* 3.2.4.1), thereby improving the directionality of the H-bonds formed between ligands and protein. The obtained alignment of the 181 ligands used for the development of the QSAR model is presented in *Figure* 55.

In general, the ligands show a common hydrogen-bonding pattern and a similar orientation of Le^x-core, but a wide variability in sidechains of the GlcNAc-moiety and in mimics of the NeuAc-moiety.

To remember is that this kind of alignment, based on the orientation of the common pharmacophoric groups, is only position and conformation dependent. For this procedure, the protein itself is not considered.
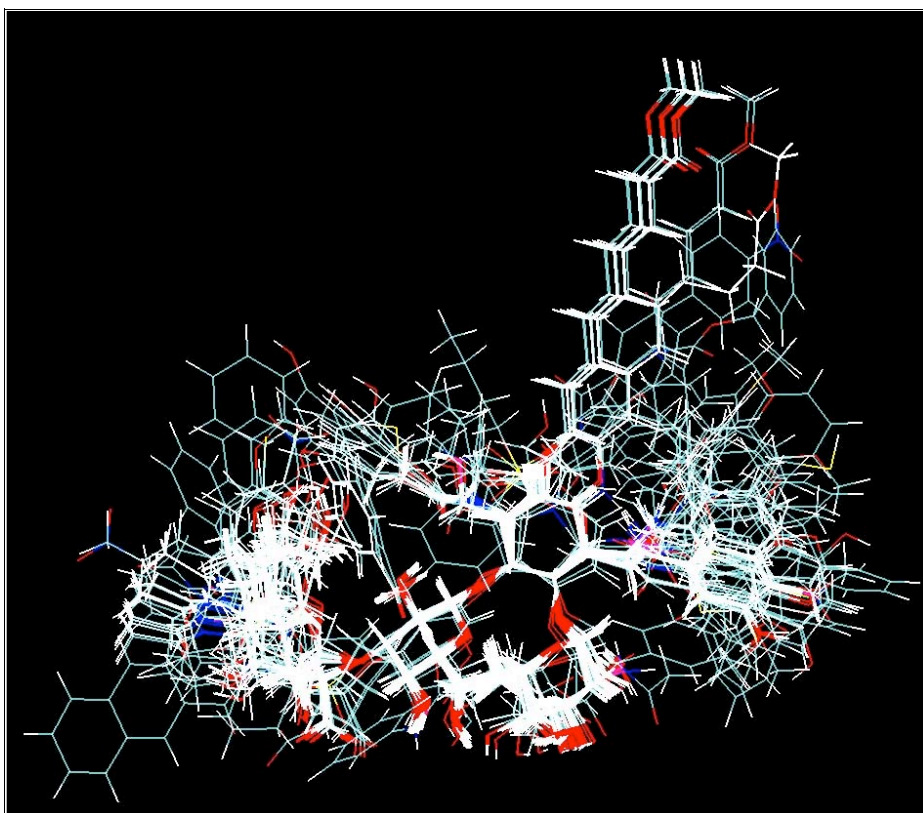
Figure 55: Pharmacophore-based alignment of the 181 ligands.

4.6.3.2 Receptor-mediated alignment

All the ligands were manually docked into the protein-binding site in an orientation similar to sLe^x as observed in the crystal structure [166]. The protein-ligand interactions were optimized through the Monte-Carlo protocol (cf. *Chapter* 3.2.5.1) to identify the energetically most favorable binding mode for the E-selectin antagonists at the receptor. By superpositioning the 200 conformations of each molecule obtained during the Boltzmann sampling over 2,000 rounds only slight differences were observed (cf. *Figure* 51). Therefore, it was decided to use only the energetically lowest conformation of each compound to establish a QSAR. The 3D superposition of the lowest-energy structures of the entire set of ligands (cf. *Figure* 52) showed large deviations between the individual ligands. The position of the sugar moieties and the orientation of the sidechains differ from each other leading to a large spatial distribution of the functional groups. This complicates the precise distribution of the atomistic properties onto the three-dimensional surface of the receptor model. However, the receptor-mediated alignment yields more "realistic" conformations and orientation of the

ligands compared to ones obtained from a pharmacophore-based approach, because it allows for the induced fit of the protein.

4.6.4 Development of a QSAR model using Quasar

The four ligand sets defined above (CM-1 align, CM-1 MC, ESP align, ESP MC, cf. *Chapter* 4.6.2) were used to generate a binding-site model based on 6D-QSAR (cf. *Chapters* 1.7.3 and 3.1.2.13). In this work, however, the fourth dimension of *Quasar* [270–272,351], the multi-representation of the conformation, was not used because every ligand of the pharmacophore and of the receptor-mediated alignment was represented only by the lowest-energy conformation obtained from the docking protocol. The reasons for that choice were explained in the previous chapter.

Different boundary conditions (*Table* 8) were tested to obtain an optimal model (high predictability, low rms deviation). In particular, models with a different number of ligands used (the ligand were divided in classes by similarity and each group was first added to the model after successful derivation of a QSAR for the previous group), different set of ligands (CM-1 align, CM-1 MC, ESP align, ESP MC), and different setting for the following parameters (dynamic surface area, hydrogen-bond function, population size, E_{solv} attenuation, number of crossovers), were generated. A detailed list of the parameter settings and the appertaining q^2, p^2, rms and the maximal deviation between the calculated and the experimental values are given in *Table* 8. All parameters not specified in this table were set as default [351]. The quality of the models was rated based on their q^2, p^2 and rms values. For the best models *scramble tests* [338] were performed to verify their selectivity towards the biological data.

Table 8: Variation of the parameters and the resulting values for q^2, p^2, rms and max. deviations in kcal/mol and in K. Q1–Q11 and Q15–Q17 were generated with the ligand set CM-1 align, Q12 and Q18–Q20 with the ligand set CM-1 MC, Q13 and Q21–Q23 with the ligand set ESP align, and Q14 and Q24–Q26 with the ligand set ESP MC. By the simulations Q12–Q14 the parameter "crop selection" [351] was set to "asymmetric 1.0". In addition, in Q14 the cross validation term was set to –2.0.

no. of the Quasar model	# total ligands (test/training)	Esolv atttenuation factor	Dynamic surface area	H-bond radii	Solv. energy and partial-charge equalization	# cross-overs	Population size	q2	p2	Training: rms / max. Deviation [kcal/mol] [in K]	Test: rms / max. Deviation [kcal/mol] [in K]
CM-1/ ESP											
1	97 (38/59)	1.0	none	gsm	none	10,000	50	0.887	0.096	0.405 / 1.206	1.114 / 2.623
2	97 (38/59)	1.0	none	gsm	none	10,000	200	0.763	0.465	0.583 / 1.429	0.843 / 1.639
3	97 (38/59)	1.0	none	gsm	none	15,000	500	0.722	0.314	0.621 / 1.521	0.946 / 2.001
4	103 (32/71)	1.0	solvent	std	none	20,000	200	0.709	0.530	0.680 / 1.555	0.736 / 1.606
5	103 (32/71)	1.0	solvent	gsm	none	25,000	200	0.763	0.533	0.617 / 1.500	0.737 / 2.147
6	103 (32/71)	1.0	solvent	ind	none	30,000	200	0.762	0.510	0.618 / 1.417	0.755 / 1.981
7	103 (32/71)	1.0	none	gsm	none	10,000	200	0.659	0.519	0.728 / 1.814	0.736 / 1.726
8	103 (32/71)	1.0	none	gsm	solvation	10,000	200	0.659	0.519	0.728 / 1.814	0.736 / 1.726
9	103 (32/71)	1.0	none	gsm	charge	10,000	200	0.659	0.519	0.728 / 1.814	0.736 / 1.726
10	103 (32/71)	1.0	none	gsm	both	10,000	200	0.659	0.519	0.728 / 1.814	0.736 / 1.726
11	103 (32/71)	1.0	void	std	none	40,000	200	0.793	0.670	0.572 / 1.647 1.7 / 15.9	0.610 / 1.264 1.9 / 7.8
12	103 (23/80)	0.3	solvent	gsm	none	17,000	200	0.786	0.467	0.568 / 1.451 1.7 / 11.1	0.744 / 1.627 2.6 / 15.3

13	99 (24/75)	0.3	void	gsm	none	37,000	200	0.888	0.667	0.384 / 1.115 0.9 / 5.8	0.533 / 1.121 1.5 / 5.9
14	103 (27/76)	0.2	void	gsm	none	17,000	200	0.790	0.367	0.511 / 1.131 1.4 / 6.0	0.863 / 2.243 3.4 / 46.1
15	181 (59/122)	0.3	none	std	none	19,000	200	0.674	0.483	0.666 / 2.052 2.1 / 33.0	0.779 / 1.599 2.8 / 14.6
16	181 (59/122)	0.2	solvent	std	none	15,000	200	0.598	0.396	0.760 / 2.176 2.7 / 41.0	0.856 / 1.789 3.4 / 20.6
17	181 (59/122)	0.2	void	std	none	32,000	200	0.699	0.470	0.647 / 2.061 2.0 / 33.5	0.791 / 1.709 2.9 17.8
18	181 (55/126)	0.27	none	gsm	none	10,000	200	0.620	0.304	0.717 / 2.688 2.4 / 100.2	0.897 / 2.446 3.7 / 65.8
19	181 (55/126)	0.23	solvent	gsm	none	10,000	200	0.570	0.324	0.776 / 2.349 2.8 / 55.5	0.888 / 2.049 3.6 / 32.7
20	181 (55/126)	0.2	void	gsm	none	25,000	200	0.638	0.346	0.699 / 2.549 2.3 / 78.7	0.865 / 2.003 3.4 / 30.2
21	181 (53/128)	0.4	none	std	none	15,000	200	0.716	0.484	0.626 / 2.194 1.9 / 42.3	0.781 / 1.737 2.8 / 18.7
22	181 (53/128)	0.4	solvent	std	none	17,000	200	0.706	0.442	0.637 / 2.262 2.0 / 47.7	0.810 / 1.779 3.0 / 20.2
23	181 (53/128)	0.3	void	std	none	5,000	200	0.529	0.350	0.803 / 2.544 3.0 / 78.1	0.876 / 2.049 3.5 / 32.8
24	181 (55/126)	0.8	none	std	none	5,000	200	0.414	0.282	0.917 / 2.754 3.8 / 112.4	0.926 / 2.462 3.9 / 67.7

25	181 (55/126)	0.7	solvent	std	none	10,000	200	0.466	0.311	0.872 / 2.747 3.5 / 111.1	0.904 / 2.411 3.7 / 63.1
26	181 (55/126)	0.6	void	std	none	5,000	200	0.401	0.309	0.929 / 2.780 3.9 / 117.6	0.909 / 2.383 3.8 / 58.9

The impact of the different parameters on the predictability of the models is briefly discussed here:

Population size

To identify the ideal size of the population, binding-site surrogates were generated with 50, 200 and 500 models respectively. Population sizes of 200 and 500 led to comparable results whereas 50 models showed small variabilities and led to poor predictions (cf. *Table* 8, Q1–Q3)[8]. Thus, for time reasons, it was advantageous to carry on with a population size of 200 models.

E_{solv} attenuation

One feature of *Quasar* [270–272,351] includes the possibility to scale the ligand desolvation energy (cf. *Chapter* 3.1.2.13). This option was particularly important in our case because the binding site is solvent exposed [166] and therefore the ligands may not be desolvated entirely upon binding. For the best models, *Quasar* [270–272,351] (cf. *Table* 8, Q15, Q21) attenuated the solvation energy by a factor of 70% approximately (cf. *Table* 8). A similar attenuation of 63% was obtained also with the most recent version of *Quasar* [270–272,351] (version 5.0), which optimizes this parameter in an unbiased fashion (→ 6D-QSAR). Interestingly, this reduction tends to have a stronger influence when CM-1 charges [327] are used. This is possibly related to the more pronounced charge separation observed when using the CM-1 method [327] (cf. *Chapter* 4.3), leading thereby to higher and more differentiated desolvation-energy values (cf. Appendix 19).

Hydrogen-bond radii

The variation of the hydrogen-bond radii of the quasi-atomistic particles showed comparable results either if this parameter was set to golden-section mean (= 0.618 • standard + 0.382 • individual) or standard (cf. *Table* 8, Q4–Q6). Slightly lower p^2 values were obtained with the parameter set to "individual". Therefore further models were generated with the parameter either set to golden section mean or standard.

Dynamic surface area

Various tests with the dynamic surface area disabled or set to, solvent or void showed no apparent trend for this option (cf. *Table* 8). For all set of ligands receptor models were generated with the three settings: none, solvent and void.

---

[8] With only 50 models, the genetic pool is not enough broad to capture all the necessary features to describe the ligand set.

However, all models other than the "none" one died during the evolutionary process. Unfortunately, this could be an artifact produced by *Quasar* [270-272,351]. Hence, the current version of the software does not allow the "solvent feature" to occupy more than the 30% of the total recetor-model surface, which is clearly too small for the E-selectin binding mode. Consequently, such models will hardly survive during the simulated evolution.

Number of crossovers

The evolution of the models is based on the *LoF* function (cf. *Chapter* 3.1.2.13). To avoid overtraining — the binding-affinity prediction of the training ligands improves whereas the difference between the predicted and the experimental binding affinity increases for the test ligands — the progression of the p^2 and of the rms value of the test set were monitored to find the optimal number of crossovers. If the p^2 and rms values remained stable the simulation was dynamically terminated. For our best models Q15 and Q21, the optimal number of crossovers was of 19,000 and 15,000 respectively (cf. *Figures* 57 and 58).

4.6.5 Analysis of the developed models

4.6.5.1 Q11-Q14

In the search for the best predictive model, new ligands were gradually added to the previous set of the ligands. The best p^2 and rms values of each ligand-set (CM-1 aligned, CM-1 MC, ESP aligned, ESP MC) were achieved with ligand-sets containing approximately 100 ligands (Q11–Q14, cf. *Figure* 56 and *Table* 8). In these test–training set arrangements (cf. *Appendix* 20) the group of the sLe^x mimics with large sidechains were excluded.

For both partial-charge models the pharmacophore aligned ligands (models: Q11 and Q13) show a more accurate prediction of the ligands than the receptor-mediated aligned ligands (Q12 and Q14) (*Figure* 56) whereas the largest difference can be found for the ligands with an IC$_{50}$ >10 mM, referred to as non-binders. The predicted binding affinity of these "non-binders" tends to higher values than experimentally measured (for a detailed discussion see below). The trend of the superior prediction for the pharmacophore-aligned ligands is more distinct when using the ESP-MNDO partial-charge method [331].
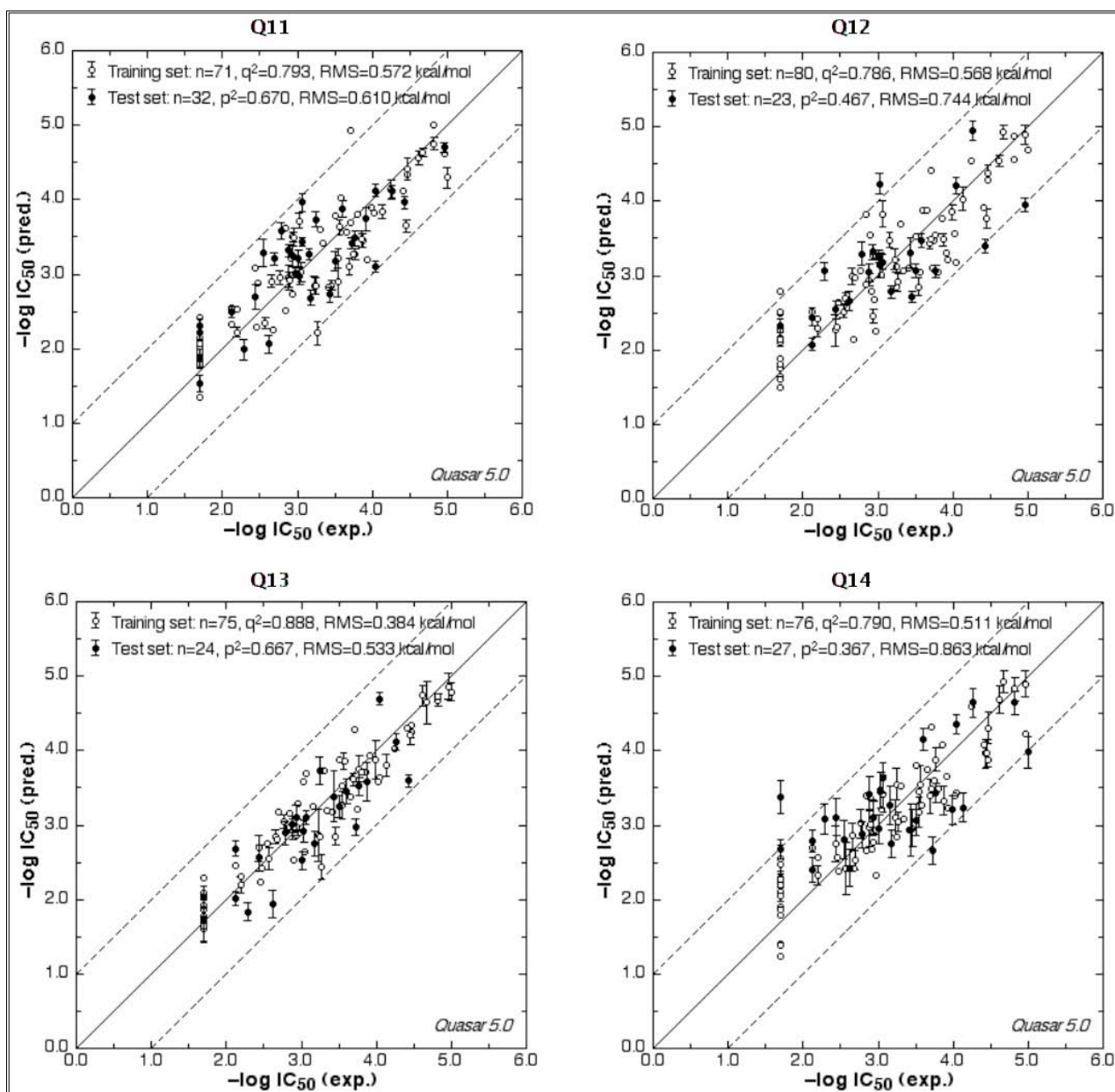
Figure 56: The results obtained for the model Q11–Q14 (cf. *Table* 8).

4.6.5.2 Q15-Q26

These models were generated by using all the 181 ligands chosen for QSAR studies (cf. *Chapter 4.6.1*). The results achieved, are somewhat less impressing (when comparing the predictive power (p^2)) than when handling with only about 100 ligands (cf. *Table* 8). This is probably due to the difficulty of the algorithm of handling ligands presenting large sidechains contributing only less to the binding

affinity[9]. However, a QSAR model considered to be good enough for our purpose (fast determination of an approximate IC_{50} of new E-selectin antagonists *in silico*) could also be established when using all the 181 compounds, and even though the results were outperformed by the model with about 100 compounds, the following models are much more complete (more structural variations are included) and with a higher predictive power for new designed ligands.

The best predictions were achieved using the pharmacophore-aligned ligands bearing either CM-1 [327] or ESP-MNDO [331] charges with Q15 and Q21 respectively (cf. *Table* 8, *Figures* 57 and 58).

---

[9] Thereby, it is assumed that the hydrophobic chains are orientated towards the solvent and do not fold back towards the protein as MD-investigations has shown.
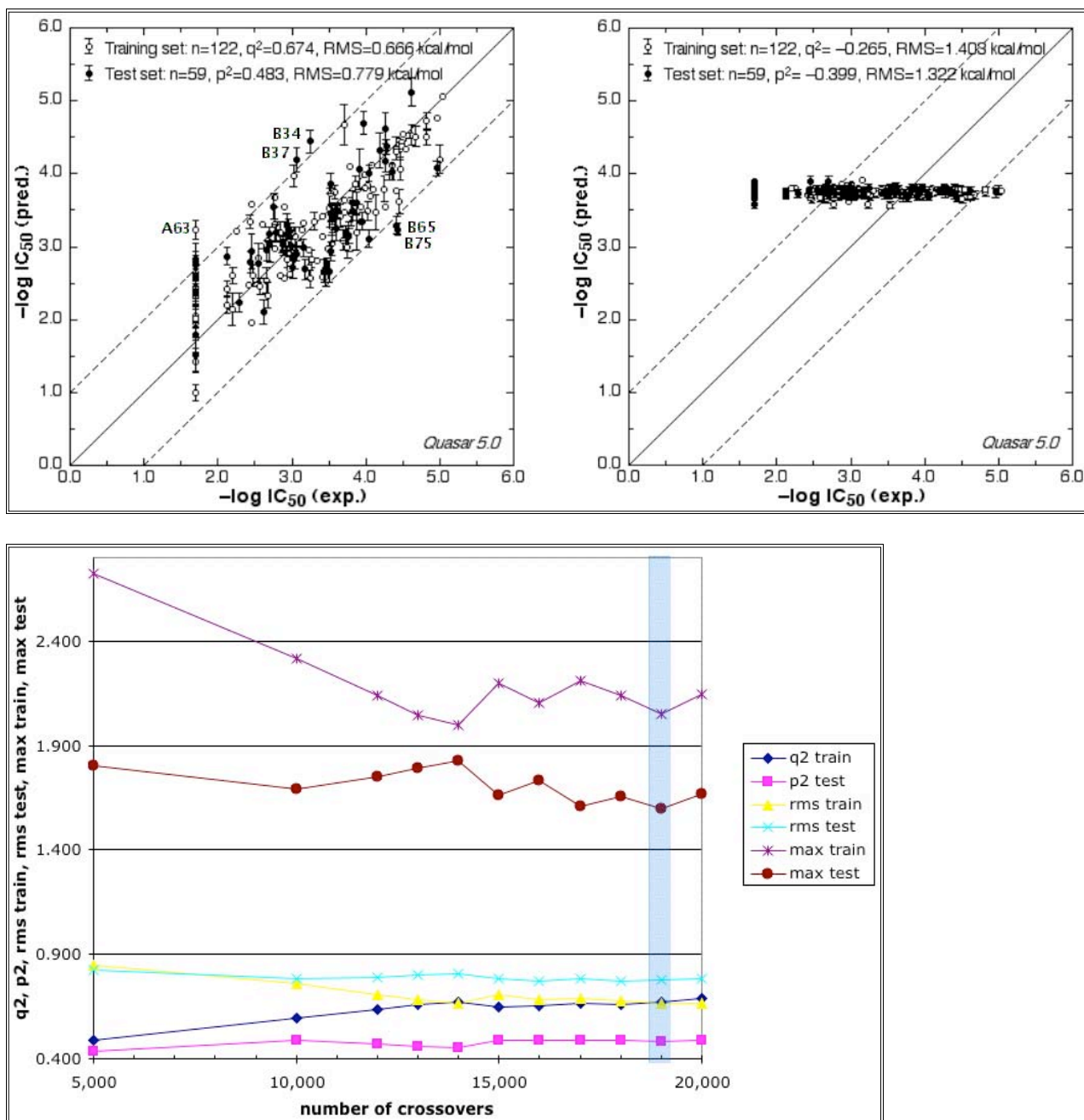
Figure 57: Results of the normal simulation, of the scramble-test simulation [355], and "time-dependent" development of the parameters (p^2, q^2, rms and maximal deviation) for model Q15.
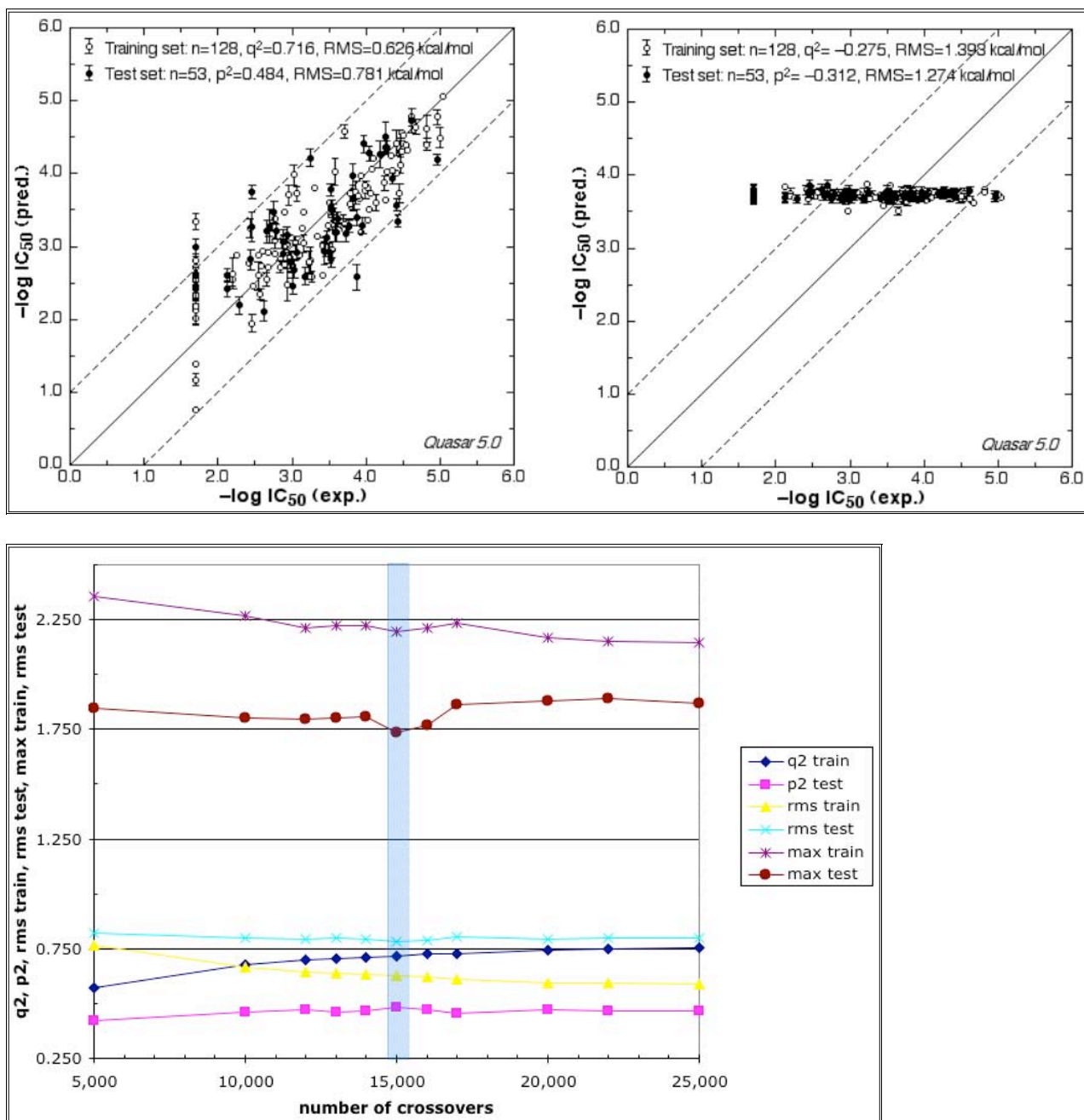
Figure 58: Results of the normal simulation, of the scramble-test simulation [355], and "time-dependent" development of the parameters (p^2, q^2, rms and maximal deviation) for model Q21.
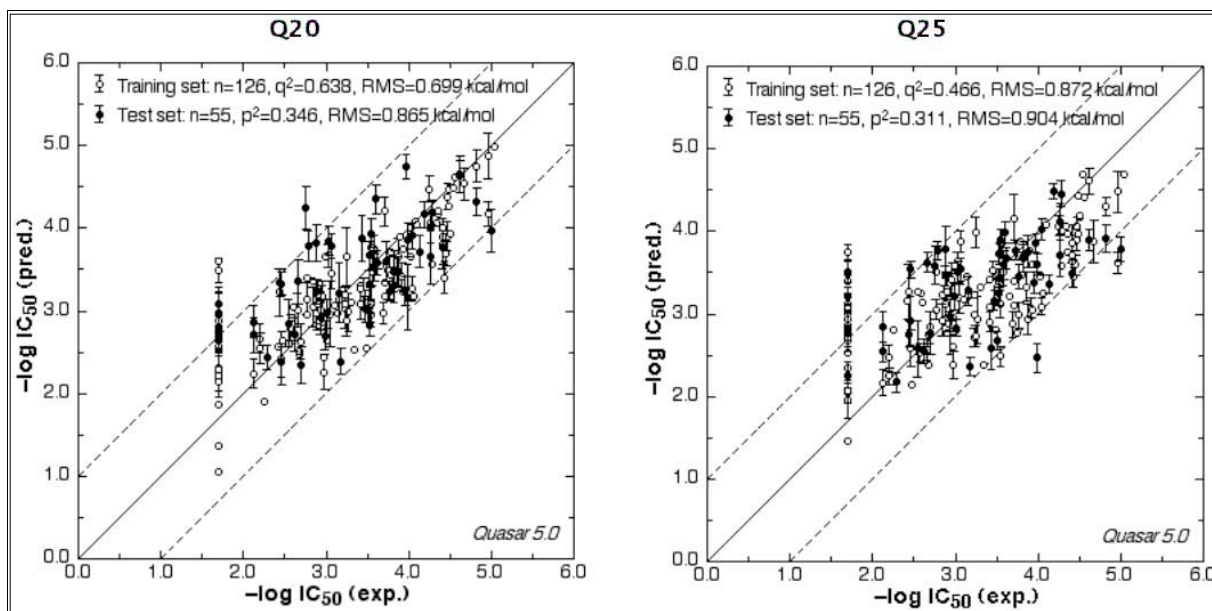
Figure 59: The best models obtained using the receptor-mediated alignment strategy and the partial-charge methods CM-1 [327] (Q20) or ESP-MNDO [331] (Q25) respectively.

On the other hand, the best models, but inferior to those aligned by pharmacophore (lower p^2, higher rms devations), developed using the receptor-mediated aligned ligands bearing either CM-1 [327] or ESP-MNDO [331] charges were achieved with Q20 and Q25 respectively (cf. *Table* 8, *Figure* 59). This observation that the pharmacophore aligned ligands yield the best predictions for all sets of ligands, might result from a better overlay of the functional groups and of the sugar-backbone (*Figure* 55) that allow a more precise mapping of the atomistic properties onto the surface than when the receptor-mediated alignment is used (*Figure* 52). Hence, the more complex alignment of the individual receptor-mediated aligned ligands might substantially affect the three-dimensional surface and the energies for the induced fit between protein and ligand. This is supported by the *Quasar* [270–272,351] study, where the induced fit was simulated upon the steric potential, leading to a significant induced fit, i.e. a snugly fitting surface. In general, the generation of good predictive receptor-models was more difficult with the receptor-mediated alignment.

The overall best model would seem to be Q15 and will be discussed more in detail below.

4.6.5.3 Prediction of the apparent binding affinity of the weak binders and of ligands presenting particularly hydrophobic moieties

An important point to observe is that in all models the predicted binding affinities of the weak binders (IC_{50} > 5 mM) and of some ligand bearing major hydrophobic moieties, tends to be stronger than experimentally measured. The cause of this effect may be associated with the fact that the binding-model surfaces of those compounds often show a substantial lipophilic fraction where hydrophobic sidechains are located (*Figure* 60). According to the X-ray study [166], those sidechains should be exposed to the solvent at the true biological receptor. Thus, the tendency of *Quasar* [270–272,351] to predict higher binding affinities for these ligands might result from artificial lipophilic interactions between the sidechains and the incorrect atomistic properties on the surface, which are probably present due to the limitation of the presence of solvent areas to an amount of maximal 30% of the total receptor-model surface, whereas in our case the binding site is exposed to the solvent for at least 60% of its surface.

In the case of weak binders however, the difficulty in predicting their binding affinity correctly could also be related to the fact that some IC_{50} values were not exactly determined but rather estimated (cf. *Chapter* 3.2.4.2). These values could lead to fuzzy input-data not really suited for the model generation. By re-determining the IC_{50}-value of some of these ligands, the range for the binding affinity may lead to a better structure-activity correlation. Then, the model would have more accurate information (training set) on the structure-activity relationship of our E-selectin antagonists.
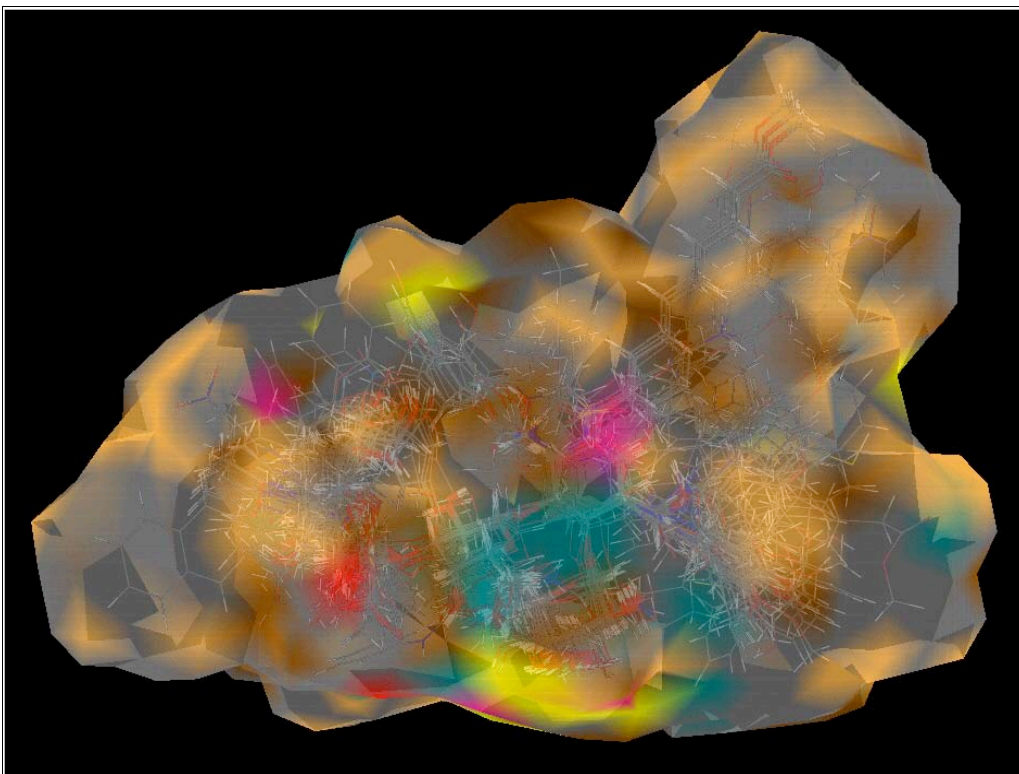
Figure 60: The three dimensional surface of model Q15 with the atomistic properties mapped onto it.

4.6.5.4 The best Quasar model: Q15

With a p^2 of 0.48±0.04 corresponding to a rms deviation of a factor 2.8 off in IC_{50}, Q15 yielded the best correlation with small standard deviations after 19,000 crossovers (cf. *Figures* 57 and 60, *Table* 8). The receptor-model was generated without allowance for a dynamic surface area and the H-bond function shift set to standard. The solvation energy was attenuated to 30%.

The prediction of the "non-binding" ligands or ligands bearing long lipophilic sidechains tends to be too strong. Ligand **A63**, a "non-binder" of the training set (cf. *Appendix* 9 and *Figure* 57), posed the major challenge for the model. Its prediction had a rms value of 2.1 kcal/mol (factor of 33 too high in IC_{50}), whereas the typical error in biological assays is varies from 0.2 kcal/mol for compounds showing an affinity in nanomolar range to 2.0 kcal/mol for compounds showing an affinity in millimolar range [215]. This large deviation is possibly resulting from the chemical structure because **A63** is the only five-membered ring with a nitrogen atom (pyrolidine ring) and a large lipophilic group. The same problem probably afflicts the ∆G prediction of the test ligands **B34** and **B37** (cf. *Appendix* 9 and *Figure* 57) with an experimental IC_{50} of 0.8 mM and 0.5 mM respectively,

which were slightly more than a factor of 10 off in IC_{50}. Their ligand structures are similar as both molecules contain a six-membered ring with a nitrogen atom (GlcNAc mimic) and large lipophilic sidechains. It is the latter that could bias the prediction as discussed above.

There are also two other ligands (**B75** and **B65**) (cf. *Appendix* 9 and *Figure* 57) with an experimental IC_{50} of approximately 40 μM that show a deviation of 12 to 15 (factor off in IC_{50}) in the prediction of the binding affinity. The two structures differ only in one N-acetyl group that seems to have no influence on the binding affinity[10]. Their structural backbone is commonly represented in the ligand set but the modified sidechains might be poorly distinguished by the binding-site model. Moreover both ligands are small in size and therefore the lower binding affinity predicted by the model could result from the fewer interactions with the surrogate compared to larger compounds.

4.6.6 Development of a QSAR-model using the Raptor technology

More recently the *Biographics Laboratory 3R* [328] developed another tool for the generation of QSAR models: *Raptor* [315,316,335]. As described in *Chapter* 3.1.1.10 section, this software is based on a different approach than *Quasar* [270–272,351]. When comparing results obtained with the two approaches for identical set of ligands, consensus scoring can be applied. In particular, *Raptor* [315,316,335] does make not use atomistic charges for model generation. Consequently, only models for two of the four set of ligands (based on the different alignment strategies: pharmacophore-based or receptor-based alignment) had to be evaluated.

The ligand-sets of Q15 and Q21 were used to perform additional QSAR simulations with *Raptor* [315,316,335] yielding the models R15 and R21 respectively. With a p^2 of 0.46 approximately the plots for R15 and R21 show a respectable correlation (*Figures* 61 and 62) and a relatively large standard deviation.

---

[10] This statement is based on the results of a functional group analysis performed with *Quasar* [270-272,351] (data not presented here).
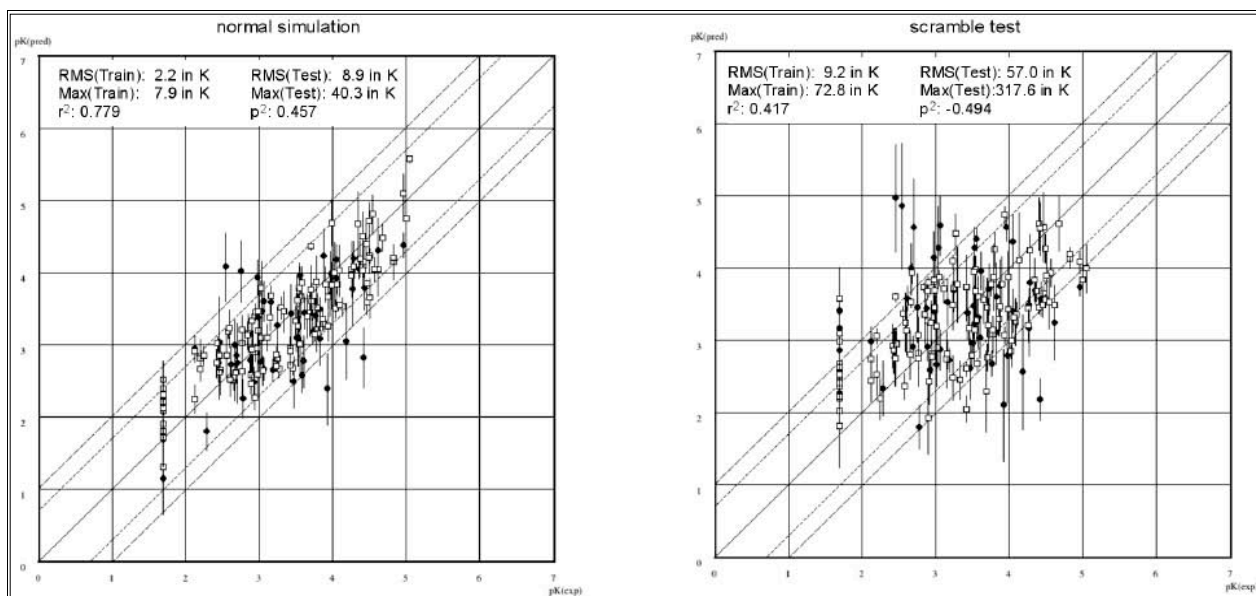
Figure 61: Results obtained for R15. On the right the results of the simulation with scrambled affinity values [338].
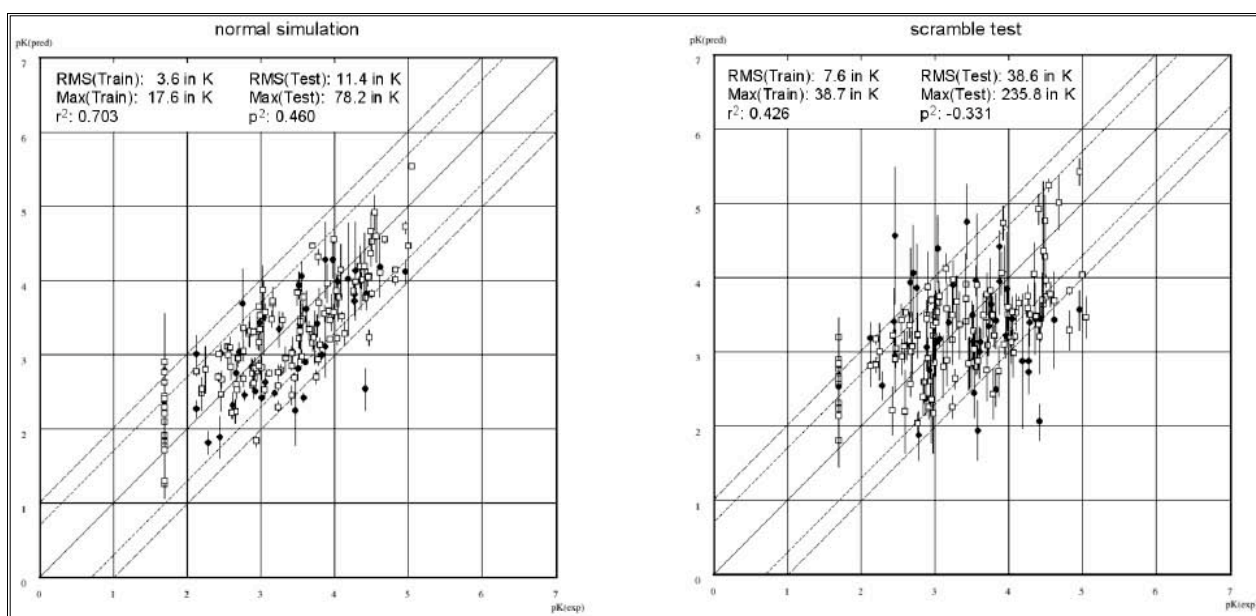
Figure 62: Results obtained for R21. On the right the results of the simulation with scrambled affinity values [338].

In contrast to *Quasar* [270–272,351], *Raptor* [315,316,335], allows for the definition a threshold value to less punish the "non-binders", when their affinity prediction is worse than the experimental value. A second model was therefore generated making use of this option with the threshold activated for $IC_{50} \geq 10$ mM to investigate its effect on the binding-affinity prediction. The plots with the predicted values versus the experimental data of these simulations (R15t and R21t) are shown in *Figure* 63. With a p^2 of 0.47 and an rms_{test} of 6.4 off in IC_{50}, R21t represents the best model (higher p^2 and smaller rms) yielded with *Raptor* [315,316,335].
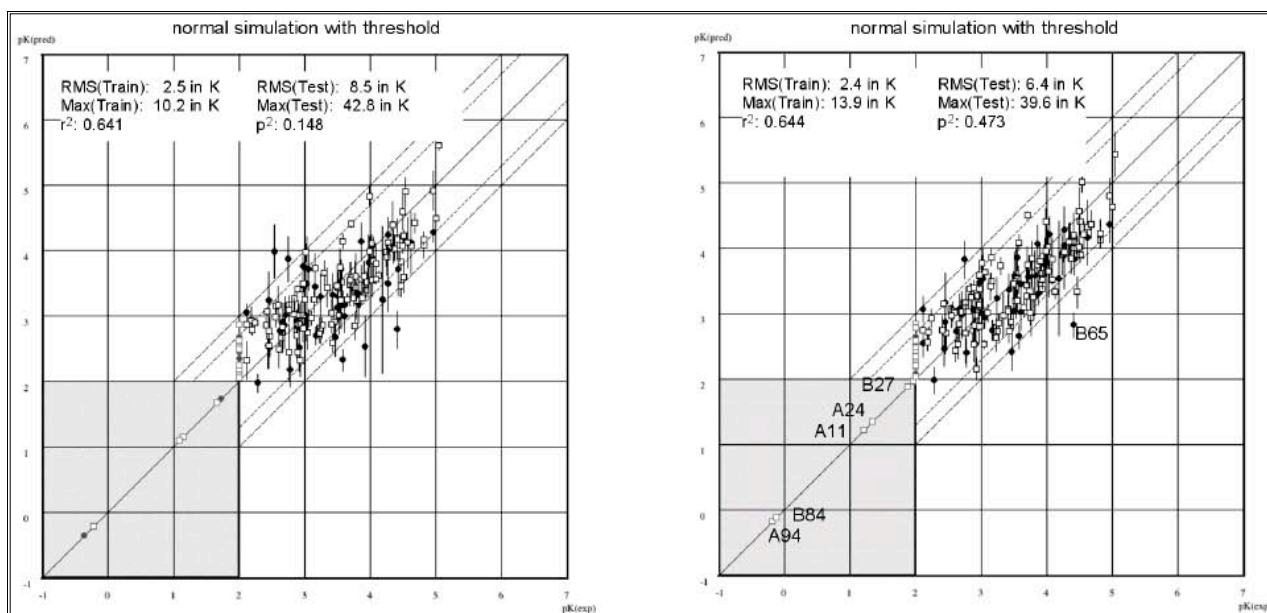
Figure 63: Results obtained for the training/test-set arrangement of R15 and R21 after allowing for a threshold (R15t left, R21t right).

The averaged model of R21t is displayed in *Figure* 64. By comparing the results achieved with and without the introduction of a threshold value, it can be noted that for the model R15 the prediction from the normal simulation (*Figure* 61, rms_{test} = 8.9 off in IC_{50}) is comparable to the one from the simulation with threshold (*Figure* 63, rms_{test} = 8.5 off in IC_{50}), whereas R21 showed a substantial improvement (rms_{test} from 11.4 to 6.4 off in IC_{50}) by enabling the threshold (*Figures* 62 and 63). The prediction of the weak binders became more accurate compared to the initial too high values and the predicted binding affinity for the ligand **B65** improved by a factor of approximately 40 in IC_{50}.
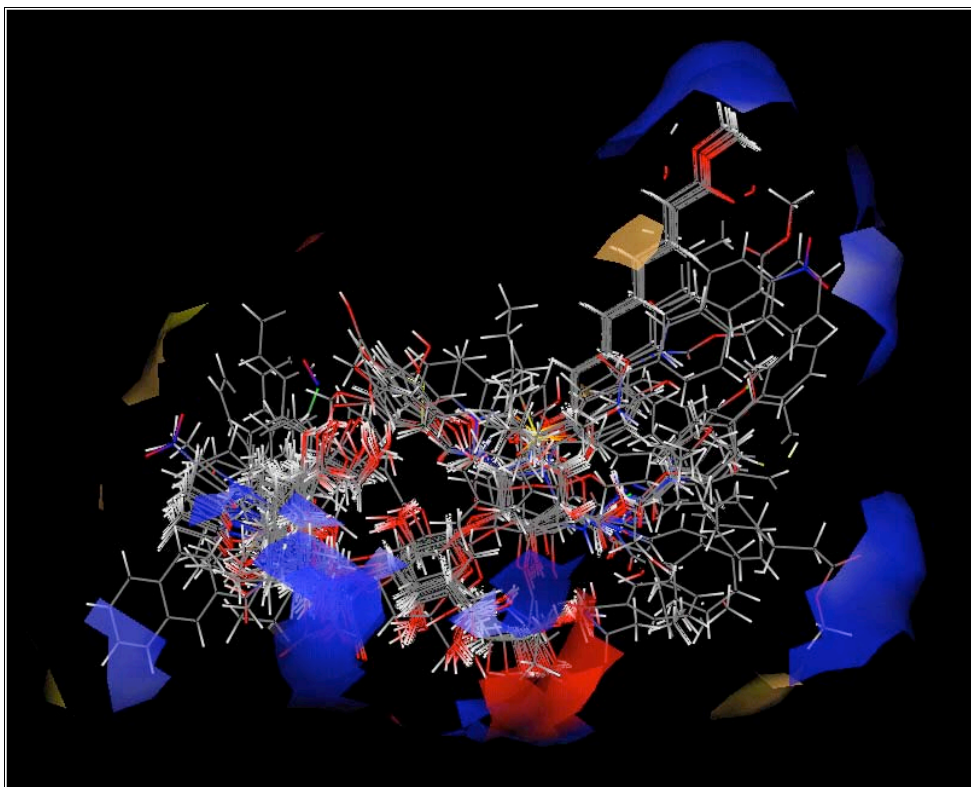
Figure 64: Illustration of all ligands with the mean model presenting the main atomistic feature of R21t. The model surface is colored as follow: hydrophobic field, beige; hydrogen-bond-donating propensity, blue; hydrogen-bond-accepting propensity, red; hydrogen-bond flip-flop, green.

4.6.7 Comparing models generated by Quasar and Raptor

Good models to predict the binding affinities of new E-selectin antagonists could be generated with both approaches, *Quasar* [270–272,351] (p^2 = 0.48, cf. *Figure* 57) and *Raptor* [315,316,335] (p^2 = 0.47, *Figure* 63). Hence, the relatively low p^2-values suffer only from the imprecise prediction of weak binders, which are not of fundamental importance for the further development of high affinity E-selectin antagonists. *Quasar* [270–272,351] required different test-training arrangements for CM-1 [327] and ESP-MNDO [331] because the calculation of the protein-ligand interaction energy considers the atomistic partial charges. In contrast here to, *Raptor* [315,316,335] neglect atomistic partial charges but instead uses the concept of hydrophobicity and calculates the ligand-receptor interactions based on hydrogen bonds, hydrophobic interactions and treats solvation effects implicitly. Therefore this approach is independent from the partial-charges models and the quality of the model depends mainly on the test training set arrangement.

Comparing the best model of each approach, Q15 achieved better rms values for

the training and the test ligands (cf. *Figure* 57), but the predicted binding affinity for the non-binders was too high (max. factor of 33.0 off in IC_{50}) favoring R21t. Regarding the maximal deviation for the test ligands, *Quasar* [270–272,351] performed well with a maximal deviation of 14.6 off in IC_{50} (ligand **B34**) compared to *Raptor* [315,316,335] with 39.6 off in IC_{50} (ligand **B65**).

In general, *Raptor* [315,316,335] predicts the weak-binders well but shows larger standard deviations and *Quasar* [270–272,351] achieves a high correlation with small standard deviations except for some of the weak binders.

4.6.8 Conclusions on the QSAR studies

The aim of these studies was to establish a structure-activity relationship based on quasi-atomistic models and using two distinct approaches (software *Quasar* [270–272,351] and *Raptor* [315,316,335]) to estimate the binding affinity of novel E-selectin antagonists. Models were derived using a total of 181 compounds for which experimental binding data (IC_{50}) are available. The best models obtained — Q15 ($p^2 = 0.48$, *Figures* 57 and 60) and R21t ($p^2 = 0.47$, *Figures* 63 and 64) — suggest that the surrogates may be used to semi-quantitative predict the binding affinity of novel compounds *in silico*.

The influence of different partial-charge models (CM–1 [327] or ESP [331]) as well as of different alignments (pharmacophore-based or receptor-mediated) on the binding-site model was studied. *Quasar* [270–272,351] yielded a model (Q15) that explained the structure-activity relationship of E-selectin antagonists with an acceptable correlation ($q^2 = 0.67 \pm 0.01$, $p^2 = 0.48 \pm 0.04$ corresponding to a factor 2.1 and 2.8 off in IC_{50}) and small standard deviations whereas the non-binding ligands posed some problems in their prediction of the binding affinity. Receptor models generated with *Raptor* [315,316,335] yielded a $q^2 = 0.64$, $p^2 = 0.47$ corresponding to a factor 2.4 and 6.4 off in IC_{50} (R21t) that could correctly identify the weak binders.

By using *Quasar* [270–272,351] and *Raptor* [315,316,335] (consensus scoring) on a large data set, the resulting receptor models are expected to predict novel E-selectin antagonists at moderate accuracy. Hence, the models derived could be used as a pre-selection tool for an *in silico* identification of high-affine E-selectin antagonists for compound synthesis, thereby reducing development time and costs as well as contributing to environmental benefits by reducing the amount of chemical waste.

4.7 Design of new ligands

Two different strategies were employed to improve the affinity of our E-selectin antagonists towards the target protein. The first one, introduced in *Chapters* 1.5.2 and 1.5.3, bases on the idea that a more rigid and/or pre-organized compound will have to pay a less entropic costs upon binding than a more flexible ligand, resulting in more affine molecule. Different studies [161,168,175] demonstrated the validity of this approach for the optimization of E-selectin antagonists. The other strategy adopted, based on the fact that by adding new functional groups to the lead compound or by replacing part of the molecule with new functional groups more complementary to the target protein, a significant increase in affinity can be reached due to a stronger ligand-protein interaction. In the next paragraphs it will be outlined how those strategies were applied to the design of novel E-selectin antagonists.

4.7.1 Gaining affinity through pre-organization of the ligand structure

Different studies [150–168] and the results achieved in this thesis (cf. *Chapter* 4.2), suggest that the conformation of sLe^x in water could be defined by a quite rigid Le^x core and a flexible neuraminic acid moiety. Starting from sLe^x (**3**) itself or from one of the most promising sLe^x-mimic developed so far (**26**) (*Figure* 65), series of compounds were designed to study the reasons of the core stability and to enhance the pre-organization of the acid moiety.
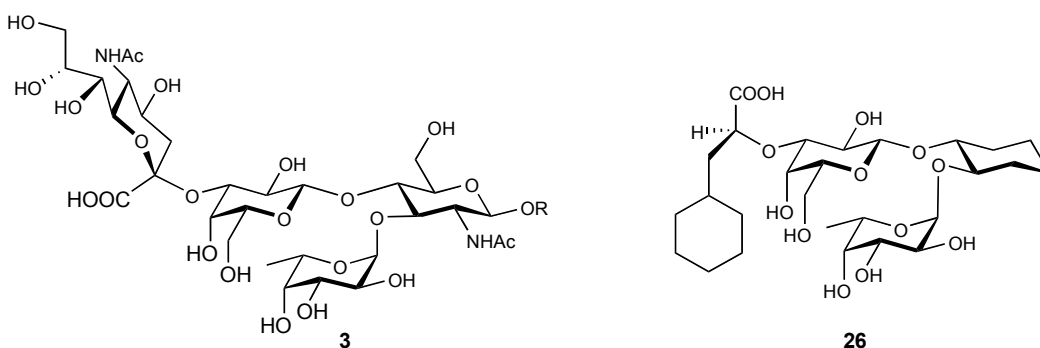
Figure 65: Chemical structure of sLex (**3**) and of compound **26**.

4.7.1.1 Investigation regarding the conformational stability of the Le^x-core

It is assumed that the exo-anomeric effect (1), the steric compression of the fucose under the β-face of the galactose moiety due to the presence of a substituent at the position 2 of the Glc*N*ac moiety (2), and the hydrophobic interaction between the methyl group of the fucose and the β-face of the galactose moiety (3) would be responsible for the stabilization of the conformation of the Le^x-core (*Figure* 66) [353].

The origin of the core stability

Three factors contribute

① Exo-anomeric effect

② Steric repulsion from substituents of Glc*N*Ac

③ Hydrophobic interactions of Fuc-C6 with the Gal β-face
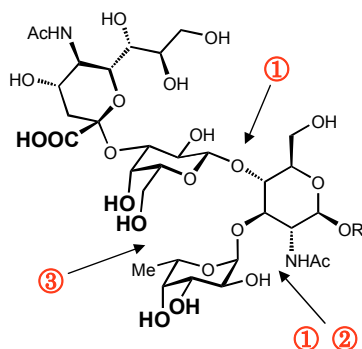
Figure 66: The three factors assumed to be responsible for the stability of the Le^x-substructure of sLe^x (**3**).

To investigate this hypothesis, three series of compounds were designed:

a) C-Glycosides

b) modified Fucose residues

c) modified Glc*N*Ac moiety

a) The C-Glycosides series

Within the series of the C-Glycosides, it could be showed that, as assumed, that the exo-anomeric effect plays a determinant role in the pre-organization of the Le^x core of sLe^x [353]. Hence, the C-glycoside analog of compound **26** (cf. *Figure* 65) resulted to be inactive, whereas **26** shows an affinity of 0.08 mM towards E-selectin. These compounds were detailed described in a recent PhD-thesis [354] and will therefore not be further discussed here.

b) The modified Fucose residues series

With the intention of studying the effect of the hydrophobic interaction between the methyl group of the fucose moiety and the β-Face of the galactose residue, compounds **77**–**80** were designed (*Figure* 67).
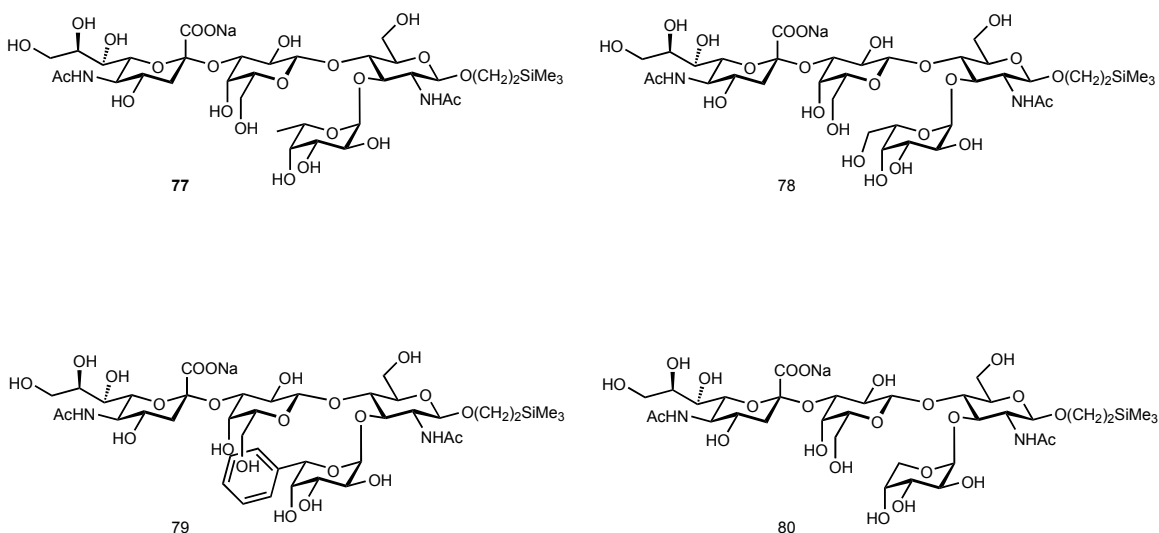
Figure 67: Chemical structure of the compounds **77**–**80**.

The four molecules were subjected to the AMBER-based protocol (cf. *Chapter* 3.2.3.2) to investigate their conformational behavior in water. As showed in *Figure* 68, compound **77** clearly the most stringent conformation distribution around the Φ_1– and Ψ_1-angles closely followed then by **78** and **79**. Compound **80**, instead, which has no substituent at the 5–equatorial position of the arabinose ring and therefore no possibility to establish a hydrophobic interaction with the β-face of the galactose moiety, occupies a much larger part of the conformational space.
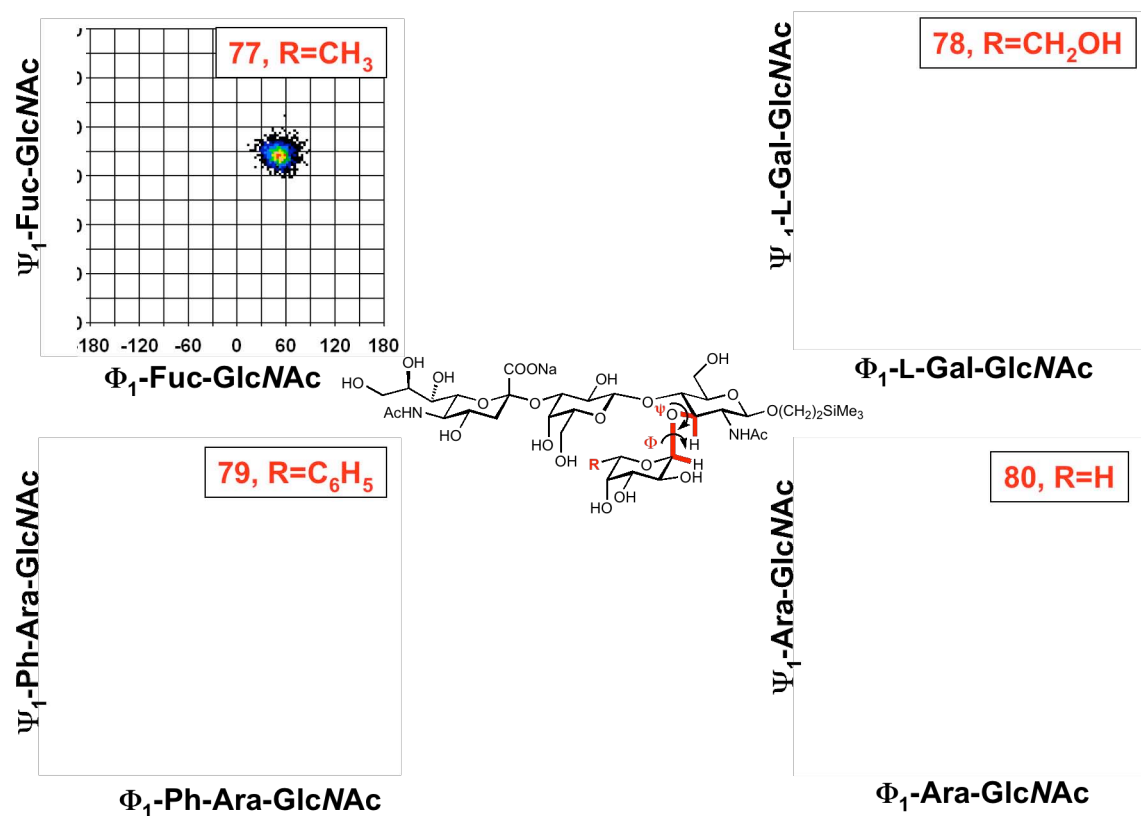
Figure 68: Conformational preferences around the Φ_1- and Ψ_1-torsion for the compounds **77–80.**

Out of those results, we predicted that, due to the different substituents at the equatorial five position of arabinose ring, which influence the strength of the hydrophobic interaction with the β-face of the galactose moiety, compounds **77–80** would have shown different affinities towards E-selectin. In particular, it was supposed that **77** would have been the most active one (in the same range of affinity as sLe^x (**3**)), followed by **78** and **79** and finally by **80**. These assumptions, based upon the results obtained by the application of one of our molecular modeling protocols, were nicely confirmed. Hence, the experimentally determined IC$_{50}$ of the compounds *in vitro,* when relative to compound sLe^x (**3**) are of the following factors higher: 1.6 for compound **78**, 2.8 for compound **79**, and higher than ten for compound **80**.

These results suggest the importance of the hydrophobic interaction between the equatorial substituent placed at the position five of the arabinose ring and the β-face of the galactose moiety. In fact, this interaction seems to be important to pre-organize the Le^x-core of sLe^x in the bioactive conformation leading thereby to ligands (**77–79**) with a higher affinity towards E-selectin than the one observed for their more freely rotating counterpart (**80**).

c) The modified GlcNAc mimics series

Since 1996 [175,176], it is known that the introduction of steric demanding substituents at the 2 position of the GlcNAc moiety or of mimics thereof led to an increased affinity of the compounds towards E-selectin. To explain this effect two different hypotheses were put forward. One, described in *Chapter* 4.4.3, bases on the possibility of a reverse docking mode. The second, on the fact, that the introduction of a substituent at the 2 position of the GlcNAc moiety would stabilize the Le^x-core of sLe^x or of the sLe^x-mimics in the bioactive conformation. In addition, the presence of a substituent at that position could also generating a benefit by liberating a larger amount of protein-bound water molecules upon binding and thereby enhancing the affinity of the ligands towards E-selectin even more. To proof this concept four compounds were designed (*Figure* 69) and submitted to the MC(JBW)/SD-protocol.
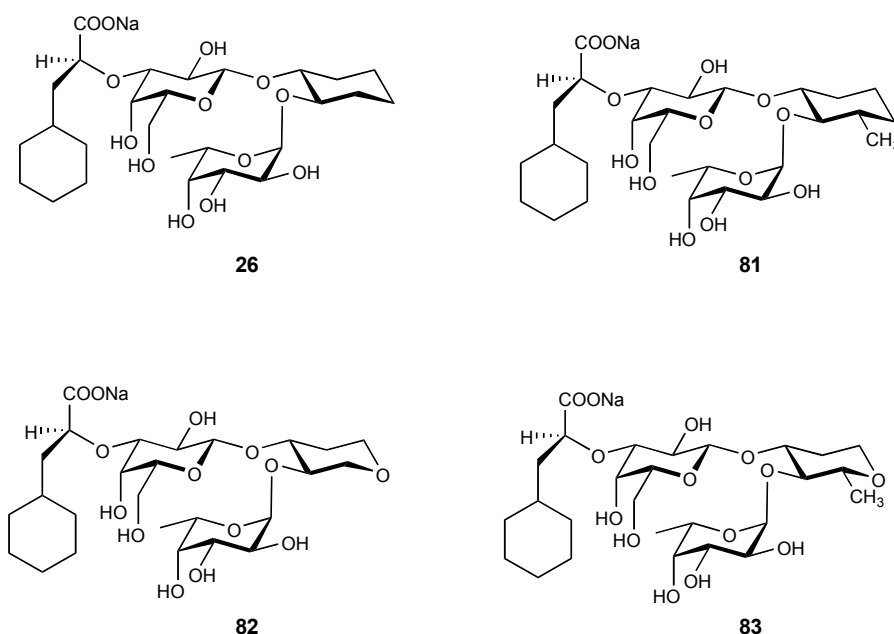
Figure 69: Chemical structure of the investigated compounds.

As it can be taken from *Figure* 70, the distribution of the conformations along the core coordinate is clearly narrowed for the compounds presenting a methyl group at the position 2 of the GlcNAc-mimic moiety when compared with the ligands missing that group. Relying on those results, we predicted that the compounds lacking a methyl group, and therefore incapable of pre-organization, would have resulted in a lower affinity towards E-selectin. Moreover, we decided to study the influence of the substitution of the ring oxygen atom in the GlcNAc-mimic moiety with a CH$_2$-group. Hence, we supposed that, by eliminating the ring oxygen

atom, we could gain even more affinity towards the target-protein due to a reduced desolvation penalty and a rigidification of the ring.
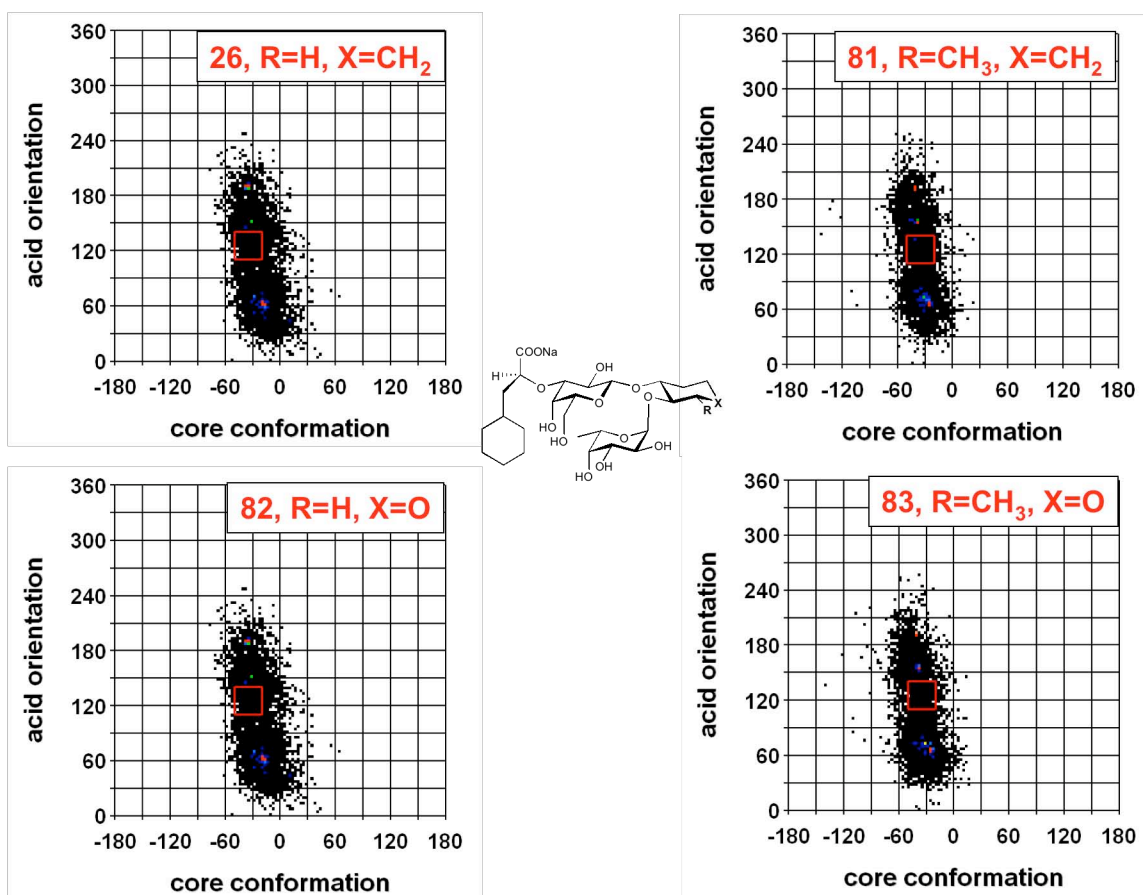
Figure 70: Conformational distribution for compounds **26**, **81**, **82**, and **83** as predicted by the MC(JBW)/SD protocol. The presence of a methyl group at the position two of the cyclohexane-ring reduces the core flexibility.

26 IC$_{50}$ = 0.08 mM

81 IC$_{50}$ = 0.018 mM

82 IC$_{50}$ = 0.23 mM

83 IC$_{50}$ = 0.04 mM
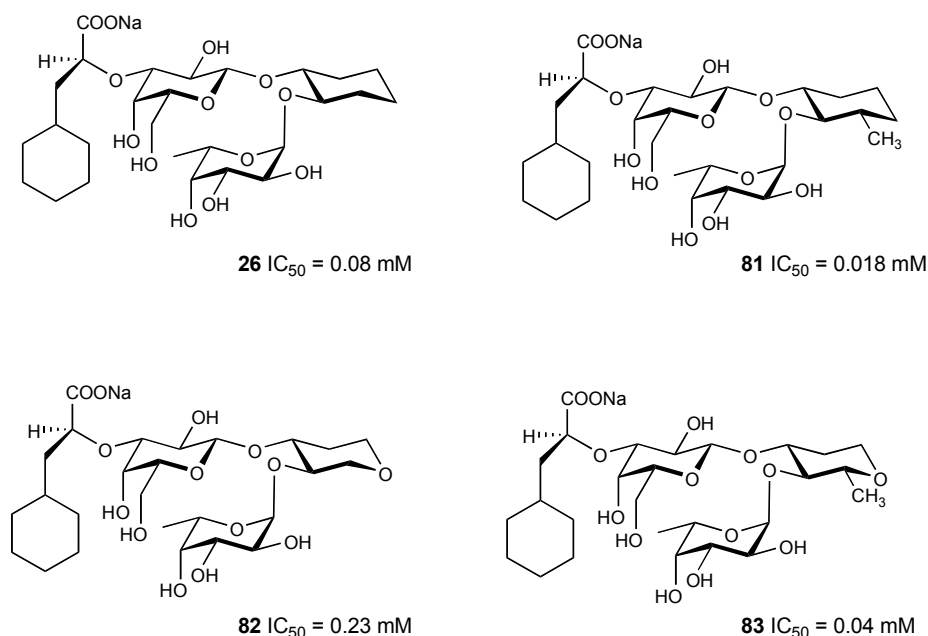
Figure 71: Chemical structure and biological activity of the investigated compounds.

As it appears from the measured IC$_{50}$ (*Figure* 71), it is interesting to conclude that all our assumption turned out to be correct. Thus, even though we cannot exclude the possibility of a reverse docking mode when more demanding substituents are attached at the 2 position of the Glc*N*Ac- or Glc*N*Ac mimic-moiety (cf. *Chapter* 4.4.3), we think to have identified the reason of the enhanced activity into a stabilization and pre-organization of the core conformation. This is probably due to the steric compression exercised by the substituents attached at the 2 position of the Glc*N*Ac- or Glc*N*Ac-mimic-moiety on the fucose residue, which is thereby fixed in its bioactive orientation under the β-face of the galactose unit. This consideration is supported by the molecular-modeling results, which show a distinct more stringent conformational freedom along the core coordinate for the compounds presenting a methyl group.

To conclude, it may be stated that the role and importance of all the three hypothesized effects (exo-anomeric effect, steric compression and hydrophobic interaction), leading to a stabilization of the core conformation of sLe^x or sLe^x-mimics could be demonstrated with the synthesis of the compounds described above. Thus, the pre-organization of the core of sLe^x or sLe^x-mimics is mandatory for achieving high affinity towards E-selectin.

4.7.1.2 Pre-organization of the acid moiety

As introduced in *Chapter* 1.5.1, one of the fundamental pharmacophoric group promoting the binding to E-selectin is the carboxylic group of the Neu*N*Ac moiety of sLe^x. From the results presented in this thesis (cf. *Chapter* 4.2) and from what is known from the literature (cf. *Chapter* 1.5.2), it is quite evident that in solution, this functional group is not pre-oriented in the bioactive conformation. It is rather supposed that a multi-conformational equilibrium exists around the Neu*N*Ac-Gal linkage. Hence, the possibility to gain affinity towards E-selectin by pre-organizing the carboxylic group in the bioactive conformation was studied. Already in 1997, Kolb and Ernst [161,168] presented two simplified sLe^x-mimics that showed a clear higher affinity towards E-selectin than sLe^x itself (*Figure* 72).

7 IC$_{50}$ = 0.35 mM   **26** IC$_{50}$ = 0.08 mM
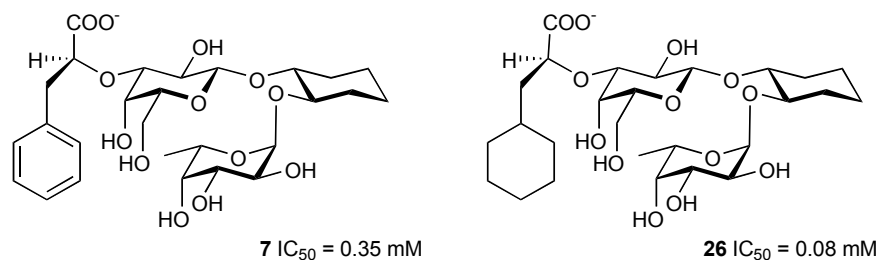
Figure 72: Chemical structure of the compounds developed by Kolb and Ernst [161,168] with the intent to pre-orient the acid moiety.

They explained this enhancement by the simplification of the ligand and by the pre-organization of the carboxylic group in the bioactive conformation. Ligand **26** served as a starting point for the design of new mimics. The underlying idea of our new development step was the introduction of steric more demanding substituents instead of the cyclohexyl-lactic acid residue of compound **26**. Thus, we supposed that this replacement would lead to another enhancement of the affinity towards E-selectin by a stronger stabilization of the bioactive conformation of the Neu*N*Ac-mimic due to the reduced rotational freedom of bulkier substituents. Nonetheless, compound **26** still shows a relatively flexibility around the torsion defining the Φ_3- and Ψ_3-angles (*Figure* 73), which was hypothesized to be reduced by the introduction of more steric demanding substituents.
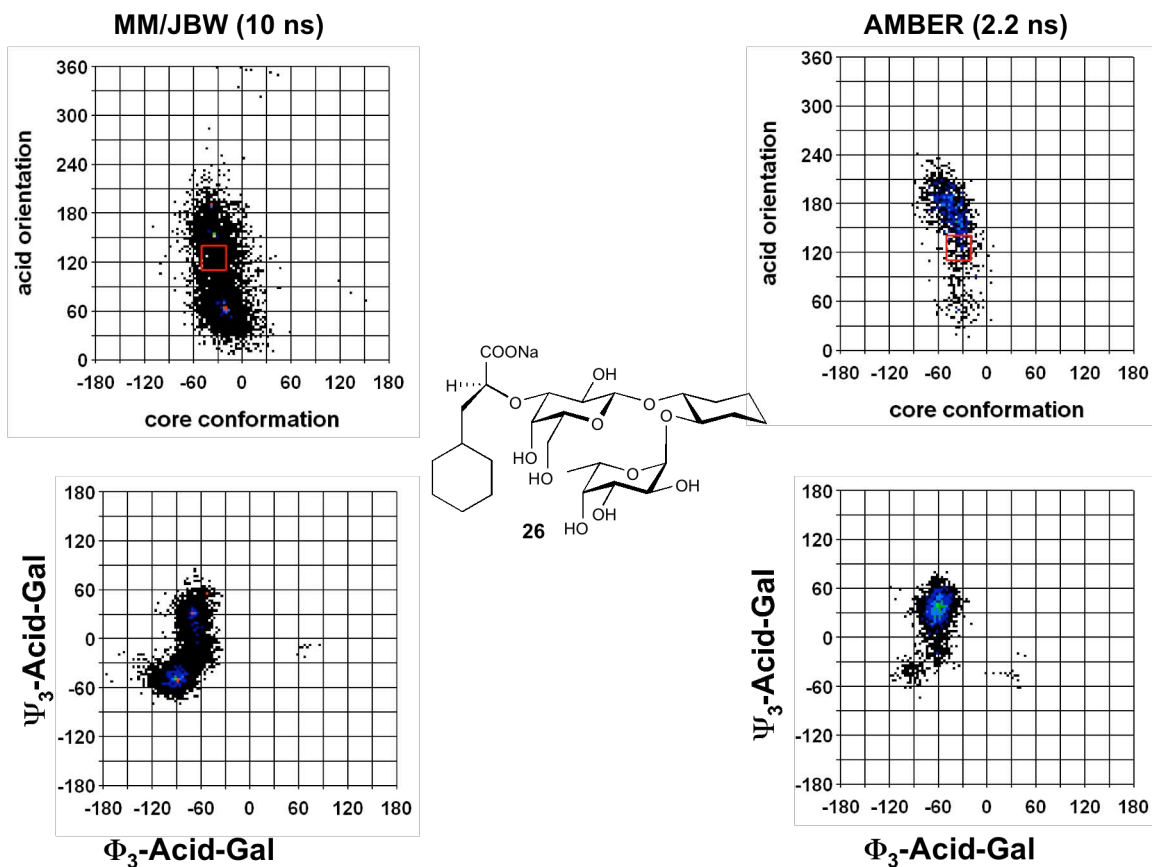
Figure 73: Conformational distribution of **26** in water as obtained with the protocols.

To support this concept, compounds **84** and **85** were designed (*Figure* 74).
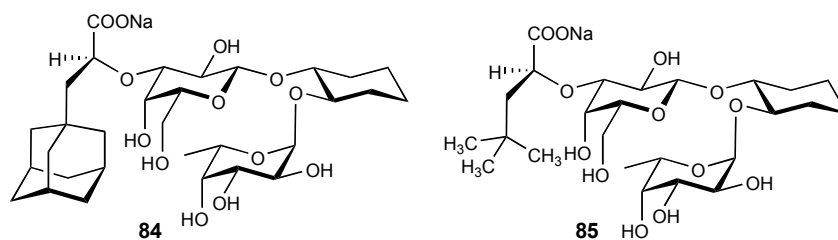
Figure 74: Chemical structure of the designed compounds.

As a Neu*N*Ac mimic, they both present groups (the adamantyl- and the tert-butyl-substituent) known to stabilize the Φ_3–torsion in the (-)-gauche conformation (the bioactive structure) and to be steric more demanding than the cyclohexyl-lactic

acid derivative. Compounds **84** and **85** were submitted to both the protocols for the analysis of the conformational preferences in water and their results compared with those for compound **26**. As it can be seen in *Figure* 75, both compounds show a more stringent conformational variability, which should results in a higher affinity towards E-selectin. The synthesis of the compounds was performed by Alexander Titz [355] and we are awaiting their biological evaluation.
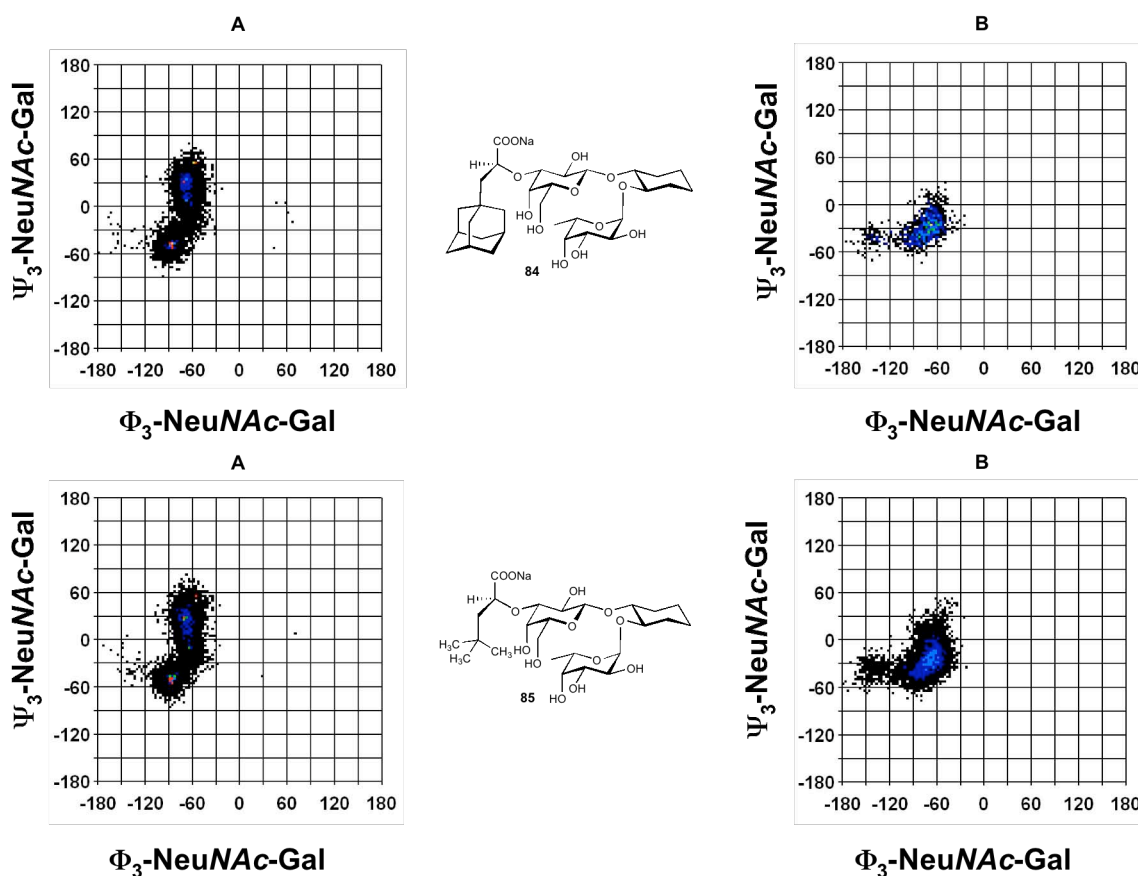
Figure 75: Simulated conformational distribution of **84** and **85** in aqueous solution as predicted by the MC(JBW)SD-protocol (**A**) and by *AMBER*-based simulation (**B**). For **84** the *AMBER*-based simulation lasted only 1.3 ns, whereas for **85,** 5 ns. The MC(JBW)SD-protocol was instead run for 10 ns for each compound.

4.7.2 Gaining affinity through additional enthalpic contributions

A more widely used method to enhance the affinity of a lead compound towards its target-protein is its modification with the replacement or addition of functional groups that increase its complementarity with the binding pocket. This strategy has gained its validity through the years [356,357] and bases on the fact that a higher complemetarity leads to higher affinity to the target-protein due to a more consistent enthalpic contribution.

In this section, our efforts directed towards the gain of affinity through new enthalpic contribution using compounds **26** as lead structure will be presented.

4.7.2.1 Modification of the sialic acid moiety

It is known that the carboxylic functionality of the Neu*N*Ac-moiety of sLe^x plays a determinant role for the binding of the ligand to E-selectin, whereas the rest of the Neu*N*Ac-moiety shows no further contacts with the protein [142,144,145]. As can be seen in *Figure* 76, also the more promising compound **26**, displays the same problem: the cyclohexyl-lactic acid moiety only stabilizes the carboxylic acid in the correct orientation but does not contribute significantly to the binding on an enthalpic level.
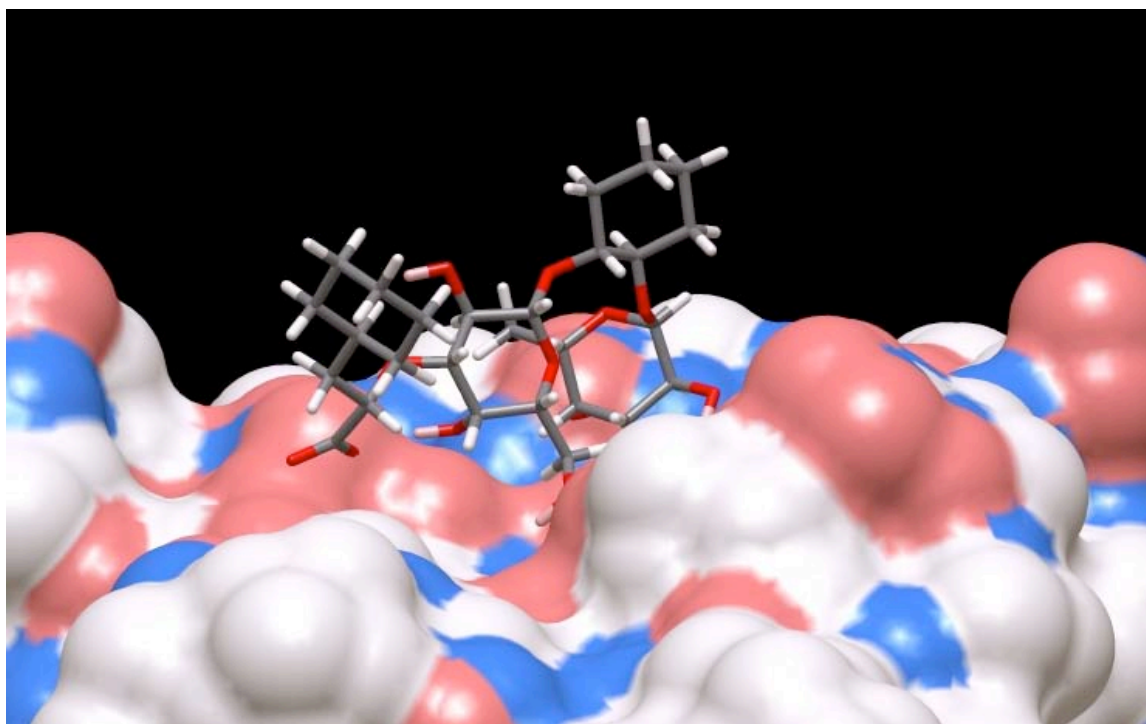
Figure 76: Binding mode of **86** as obtained from the MC-protocol (cf. *Chapter* 3.2.5. The cyclohexyl-lactic acid moiety is oriented toward the solvent.

Thus, it was decided to design a new series of compounds bearing functional groups introduced at the β-position with respect to the acid that would maintain the pre-organization of the acid moiety but also interact with the protein. From a synthetic point of view, a very attractive strategy would considered to be the possibility to introduce an azide in the β-position with respect to the acid giving rise to compound **86**. Hence, this is likely to lead to a triazol-based library of Neu*N*Ac-moiety mimics over the 1,3-cycloaddition of acetylenes to **86** (*Scheme 1*).
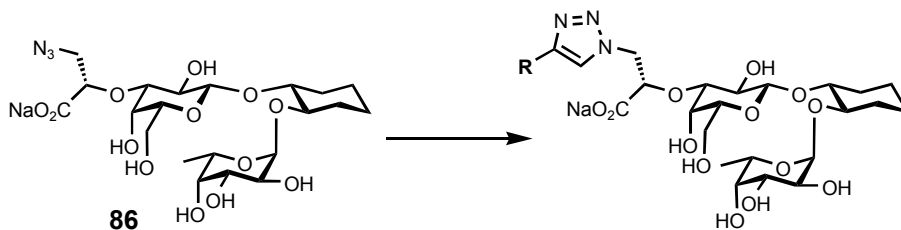
Scheme 1: Chemical structure of **86** and of the resulting triazol-based library after 1,3-cycloaddition of acetylenes.

By looking at the feature most present (H-Bond donors, H-acceptor, hydrophobic side chains,…) in the region binding the acid moiety of the sLe^x-mimics, the presence of many hydrophobic amino acids was noticed (*Figure* 77). Hence, a small library of hydrophobic substituted acetylenes to be added to compound **86** was planned for synthesis. The resulting compounds are expected to show an increased the affinity towards E-selectin because of new hydrophobic contacts with the protein.
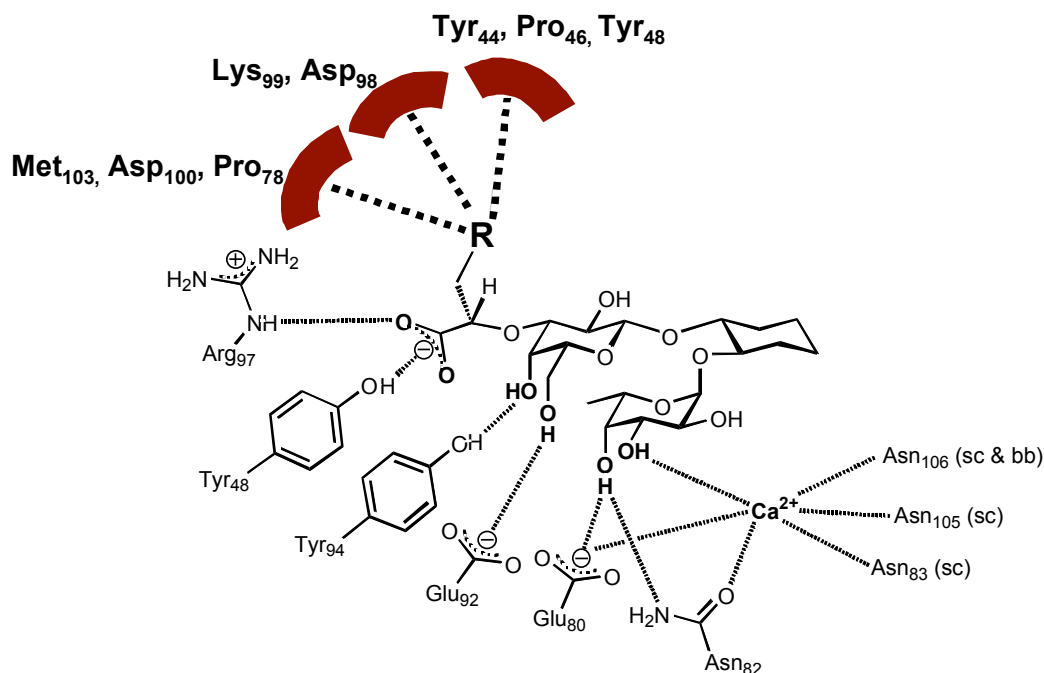
Figure 77: Hypothetical binding mode of the triazol-based compound library. **R** = triazol-based library.

Due to the fact that the pre-orientation of the acid functionality in the bioactive conformation remains an important issue, it became necessary to investigate the conformational behavior of compound **87** in solution to prove that the substitution of the cyclohexyl-rest with a triazol-based substituent would not lead to inactive compounds. The choice of ligand **87** for the conformational study was driven by the consideration that this compound would probably correctly represent the conformational preferences of the whole planned library. As it is shown in *Figure* 78, the conformational distribution obtained with both protocols observed for **87** is very similar to the one observed for **26** (cf. *Figure* 73) implying that the two compounds should show also comparable affinity towards E-selectin and suggesting higher affinities for the library compounds.

The synthesis of the compounds was performed by Alexander Titz [355] and their biological evaluation is ongoing.
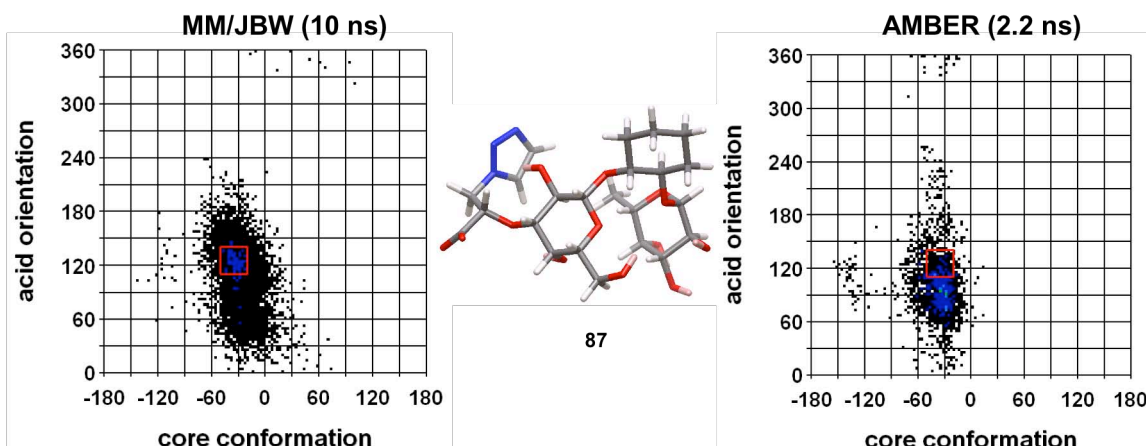
Figure 78: Simulated conformational variability of compound **87** in solution.

4.7.2.2 Modification of the galactose unity of compound 26

Only little success was achieved when it was tried to substitute or modify the galactose unity of sLe^x in the search for more potent ligands (cf. *Chapter* 1.6.2). In fact, this unit is known to play a determinant role for the interaction between sLe^x and E-Selectin [143]. Thus, the galactose unity contributes directly to the binding over two H-bonds between 4–OH of Gal and Tyr94, and between 6–OH of Gal and Glu92, but also indirectly by functioning like a three dimensional linker between the Neu*N*Ac- and the fucose-moieties. A complete substitution of the moiety is therefore difficult. However, by analyzing the protein surface contouring the galactose moiety (*Figure* 79), it appears that this moiety can still be optimized. Hence, two hydrophobic pockets located in front of the position 2 and 6 of the galactose unit, but are actually not unoccupied by the ligand. The design of new ligands based on compound **26** but reaching these two pockets, was therefore planned. To achieve this goal it was decided to elongate the chains at the position 2 or 6 of the galactose unity.
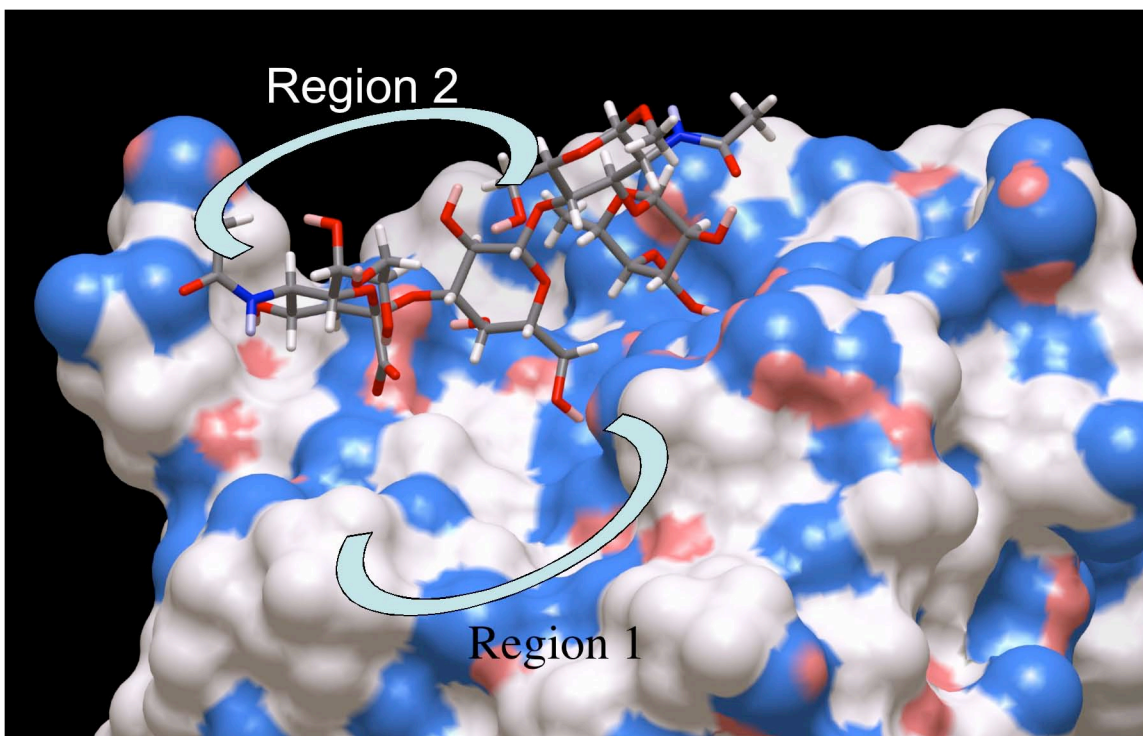
Figure 79: sLe^x (**3**) bounded to E-selectin [166]. Two hydrophobic pockets are visible and are here highlighted as *Region 1* and *Region 2.*

*a) Modification of the position 2 of the Galactose unity of compound **26***

This project was developed together with Lionel Tschopp [358], who synthesized the compounds presented in the following. The strategy, chosen for reaching the region referred as two in *Figure* 79, was, as it was already the case for the modification of the acid moiety (cf. *Chapter* 4.7.2), the introduction of an azide. This time it was decided to substitute the hydroxyl group at the 2 position of the galactose ring. The azide moiety introduced, would then be coupled over a 1,3-cycloaddition reaction with a library of acetylenes (*Scheme* 2) in the last step of the synthesis, thereby giving rise to several compounds bearing different substiutents at the position 2 of the galactose unit.
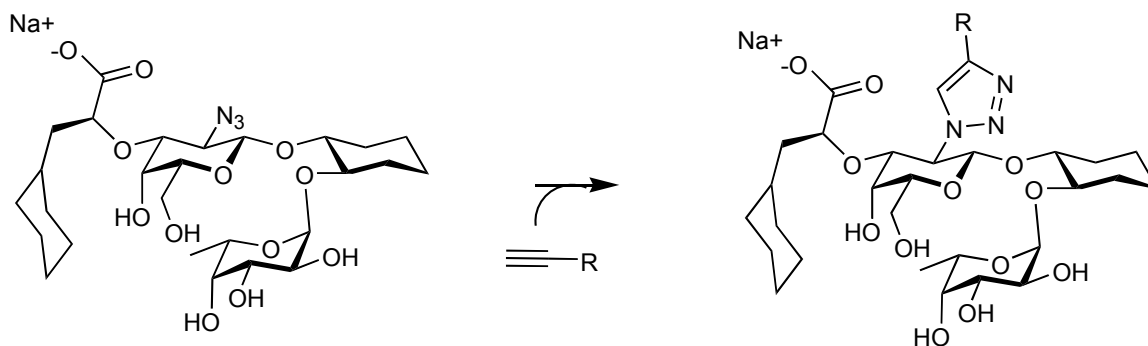
Scheme 2: Last step of the synthesis of the compounds bearing a triazol-based substituent at the position 2 of the galactose moiety.

Due to the characteristics of the binding site (cf. *Figure* 79, *Region* 2), particularly long and hydrophobic substituted acetylenes were chosen. The synthesized compounds, presently under biological evaluation, are presented in *Figure* 80. The predicted binding mode of compound **88** is shown in *Figure* 81. The orientation, which the substituent at the position 2 of the galactose unity adopts relative to the cyclohexan ring, is interesting. As it can be observed from *Figure* 81 for molecule **88**, the two groups are found to stand parallel to each other with a distance between them of about Ångstroms four. Hence, this finding suggests the presence of an intramolecular interaction between the atoms of the cyclohexan ring and of the triazol ring at the position 2 of the galactose unit, stabilizing the molecule in the bioactive conformation. A similar arrangement was found also for compound **74** (cf. *Figure* 53) in the docking experiments. In this case however, the benzyl moiety is too short-reaching to be able to interact with the hydrophobic binding pocket present on the surface of E-selectin. Though, its affinity towards E-selectin is ten times higher than for its debenzylated counterpart, thereby partially confirming the possibility of a stabilization effect of the bioactive conformation of the mimic. In conclusion, the designed compounds could raise the affinity towards E-selectin to another level by combining the stabilization of the bioactive conformation with new enthalpic contributions obtained by the newly introduced substituent at the position 2 of the galactose moiety.
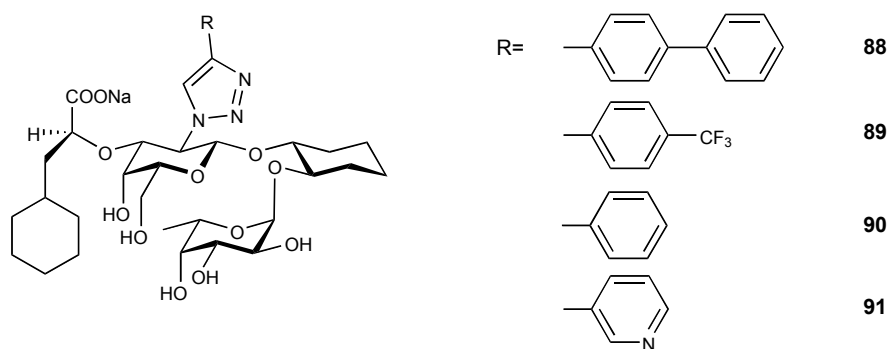
Figure 80: The compounds synthesized so far bearing a modified substituent at the 2 position of the galactose unit.
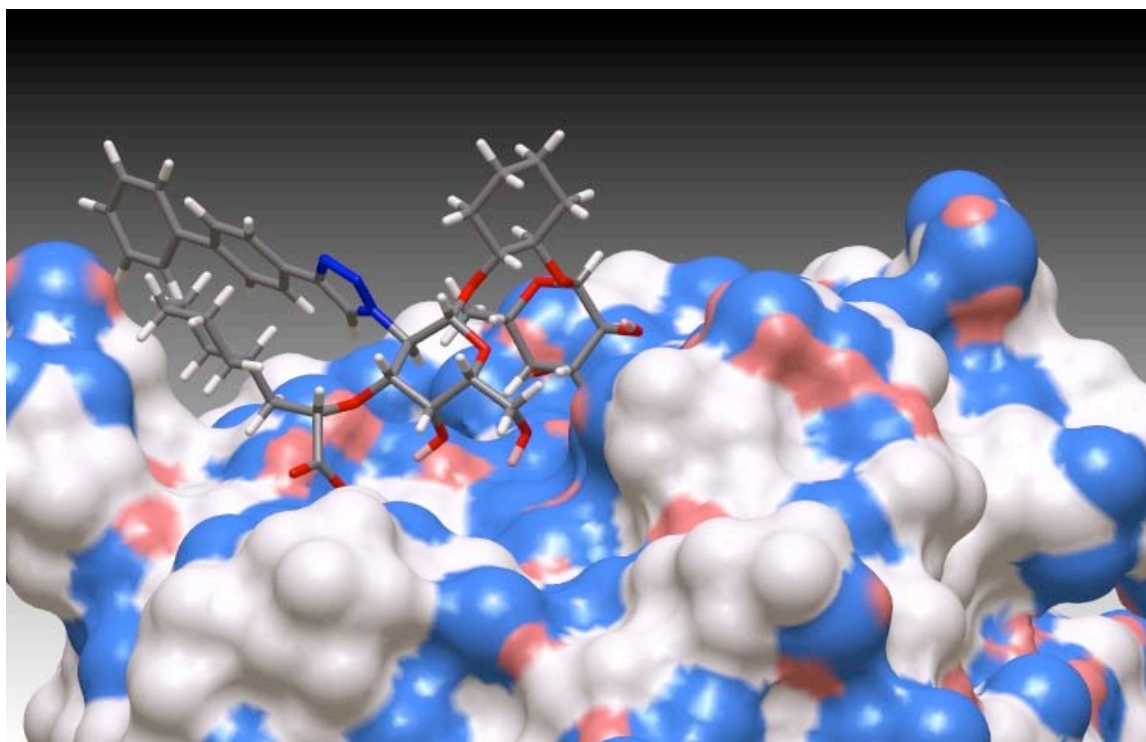
Figure 81: Compound **88** docked in the E-selectin binding site. A possible stabilization of the acid moiety in the bioactive conformation may be possible through the interaction between the atoms of the triazol- and the cyclohexane-ring.

*b) Modification of the position 6 of the Galactose unity of compound **26***

A shallow hydrophobic pocket (cf. *Figure* 79: *Region* 1) opens itself in front of position 6 of the galactose unity of sLe^x. Unfortunately, a substitution of the hydroxyl with an azide is not possible because, as it was shown by the experiments of Bänteli and Ernst [359], that any modification of the hydroxyl group leads to inactive compounds. Therefore, it was decided to keep the hydroxyl group but to elongate the chain by modifying the galactose to a heptose. The adjacent carbon atom should then serve as a starting point for the development of modified mimics that could fill the hydrophobic pocket formed by the amino acids Lys111-113, Pro46, Tyr44. At this occasion we made use of the *de novo* design software *Allegrow* [318,319] to obtain insight on the substitution pattern that could be added to the heptose moiety. The technical details are described in the methods section (cf. *Chapter* 3.2.8). Between the several thousand compounds that were generated by the first "*Allegrow*-round", some recurrent patterns were identified. In particular, indols and byphenyl moieties were frequently found to be connected over an amide bond to the heptose. After few more rounds of compounds generation by *Allegrow* [318,319] and optimization of the output towards chemical more feasible molecules by the user, the ligands presented in *Figure* 82 were chosen for synthesis and biological evaluation. All the designed ligands, with the exception of compound **92**, which has already been synthesized and biologically tested, are still under development. Also because the results obtained so far are controversial (an IC$_{50}$ could not be determined) and need to be further elucidated, a discussion on these compounds is not possible at the moment. Details will be discussed in the thesis of Daniel Schwizer [360], in which a detailed description of these new ligands, their synthesis and biological evaluation will be found.
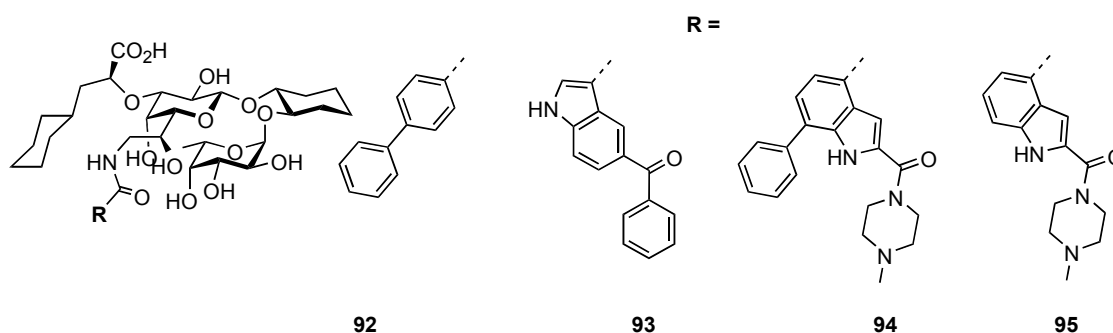
Figure 82: The compounds selected from the *de novo* design approach.

4.7.2.3 Conclusion on gaining affinity through new enthalpic contributions

Three different approaches to obtain higher affinity through new enthalpic contributions have been presented. Unfortunately, to date no biological evaluation of the compounds has been performed (with the exception of compound **92**) and it is therefore difficult to make any conclusion on the real benefit of the planned modifications. In principle, all the designed compounds should show an increased affinity towards E-selectin due to hydrophobic interactions with the protein surface. It is known [361] that this kind of interactions is of particular importance for a strong binding of ligands to protein. Those are, however, completely missing in the case of sugar-lectin contacts. Hence, in our opinion, the introduction of hydrophobic moieties should bring a strong enhancement (at least a factor 100) of the ligand's affinity towards E-selctin. Next to the enthalpic considerations, their introduction could also have an effect on the entropic level by reducing the desolvation costs to be paid upon binding and by the augmented quantity of water molecules liberated from the protein surface after ligand binding.

5 Conclusions and Outlook

The main goal of this work that was to establish a set of tools that would permit the analysis, the design and the evaluation of E-selectin antagonists *in silico* was achieved. Hence, different protocols were developed and validated (cf. *Chapters* 3.2 and 4). In particular, a new procedure for the analysis of the conformational preferences of E-selectin antagonists was established (cf. *Chapter* 4.1) and the results compared to the ones obtained with the MC(JBW)/SD approach, which had already demonstrated its validity in the past [161,168]. Thus, the comparison between the two protocols shows that the two methods predict different conformational preferences for the orientation of the sialic acid moiety of sLe^x (torsions Φ_3 and Ψ_3 in *Figure* 83). This result reflects the contrasting opinions existing on the conformation adopted by sLe^x in solution [150–168].
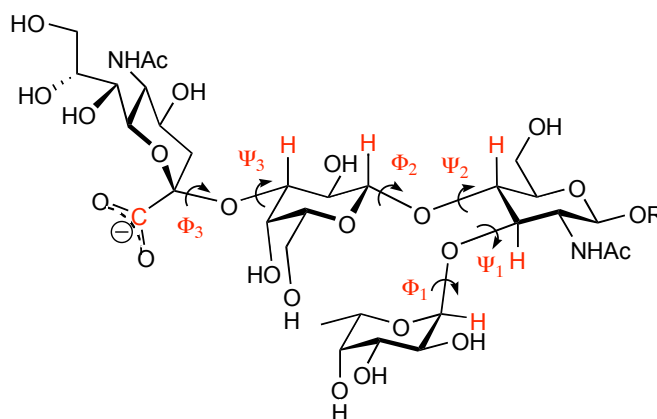
Figure 83: sLe^x (**3**) and the Φ, Ψ convention for the definition of the glycosidic torsions.

A more detailed analysis (^{3}J-coupling calculations, cf. *Chapter* 4.2) revealed that probably, both approaches deliver only a partially correct solution and that in reality, in solution, sLe^x (**3**) exists as a mixture of low energy conformers and not as supposed to date [150–154,161–163] as a population of a single conformer. Obviously, this finding is important. Hence, as a consequence, the idea to design mimics with a pre-organized structure, and in particular a pre-oriented acid moiety gains more importance. Thus, as seen in *Chapter* 1.7.1 (*Figure* 12), small differences in $\Delta G°$ have a substantial influence on the population of conformers. In our particular case, where the bioactive conformation doesn't coincide with the global energy minimum in solution, this leads to important consequences in respect to the binding affinity.

Furthermore, a docking routine was established and the impact of different partial-charge methods and of explicit solvation on the binding mode was studied (cf. *Chapter* 4.3 and 4.4). The obtained results, served as an input for molecular-dynamics simulations and QSAR-studies.

The MD simulations gave an insight into the dynamic character of the protein-ligand interactions. In particular, the observations done in an atomic-force microscopy study [350], describing the interactions between the carboxylic group of sLe^x and Arg97, and between the 3– and 4–hydroxyls of Fuc and the calcium ion as the two main energy barriers for the dissociation process of the protein-carbohydrate complex, could be confirmed by our MD-investigations. Thus, these two contacts always lasted longer than any other in the MD simulation.

The successful derivation of QSAR-models with *Quasar* [270–272,351] and *Raptor* [315,316,335] (cf. *Chapter* 4.6) gives us the possibility of a semi-quantitative *in silico* estimation of the binding affinity for the ligands that will be designed in the future.

In addition, the developed protocols and models were applied for the development of new ligands. Unfortunately, to date, only few biological data is available to evaluate our design strategies. However, the impact of the ligand's

pre-organization on the binding affinity could be established at least for the Le^x-substructure of sLe^x. Hence, the importance of the exo-anomeric effect, of the steric compression, and of the hydrophobic interaction between the methyl group of fucose and the β-face of galactose was clearly demonstrated (cf. *Chapter* 4.7.1).

In my opinion, in the future, our efforts should be directed towards the synthesis of compounds featuring more than one of the modifications demonstrating to positively influence the binding activity and of more drug-like compounds. In fact, all ligands synthesized so far do neither respect the Lipinski rule of five [362] nor have the characteristics of a drug [363]. Moreover, it would be interesting to design conformationally restricted compounds to study the entropic aspects deriving from the flexibility of the Φ_3- and Ψ_3-torsions.

Regarding the molecular-modeling protocols developed, further improvements are possible. Thus, a newer set of GLYCAM parameters [344] that could be used in conjunction with the AMBER-based approach to study the conformational preference of the E-selectin antagonists has been released. Furthermore, once the biological data of the newly designed ligands would be available, it should be integrated in the QSAR-models to ameliorate their predictability. Finally, further MD-studies would be necessary to investigate in detail the role played by the solvent in the carbohydrate-lectin interactions.

6 Appendix

Appendix 1: Listing of the ligands that were submitted to a conformational search.

3	4	5	6	7	26	27	36
65 (monomer)	72	73	74	77	78	79	80
81	82	83	84	85	87	A41	A77
A83	B45	B57	B68	B71			

Appendix 2: Command-files for the MC(JBW)/SD protocol (cf. *Chapters 1.5.3* and *3.2.3.1*). The first one refers to the CS-step, the second one to the JBW/SD-step.

Command file for the conformational search

ligand.dat
ligand.out

MMOD	0	1						
SOLV	3	1	0	0	0	0	0	0
EXNB	0	0	0	0	0	0	0	0
FFLD	3	1	0	0	1	0	0	0
SPMC	52000	0	24	0	0	0		
MCNV	2	26	0	0				
MCSS	2	0	1	0	100			
MCOP	1	1						
DEMX	0	200	0	0	20	40		
COMP	1	2	3	4				
COMP	5	7	8	10				
COMP	12	15	16	18				
COMP	22	23	25	26				
COMP	27	30	31	32				
COMP	33	34	35	36				
COMP	37	39	43	46				
COMP	47	48	49	50				
COMP	52	56	57	58				
COMP	59	60	61	62				
COMP	68	72	73	74				
COMP	75	76	77	78				
COMP	80	81	82	83				
COMP	84	85	86	87				
COMP	88	89	90	91				
COMP	92	93	94	95				
COMP	97	98	99	110				
COMP	111	0	0	0				
READ								
CHIG	1	2	4	5				
CHIG	22	33	34	35				
CHIG	37	57	59	60				
CHIG	61	62	78	82				
CHIG	83	84	85	86				
CHIG	95	0	0	0				
TORC	16	15	10	5	0	90		
TORC	32	31	110	111	0	90		
TORC	50	49	39	34	0	90		
TORS	1	7	2	8	0	180		
TORS	2	31	4	22	0	180		
TORS	5	10	8	84	0	180		
TORS	22	23	22	78	0	180		
TORS	26	78	26	27	0	180		
TORS	33	56	34	39	0	180		
TORS	35	48	37	43	0	180		
TORS	43	46	56	57	0	180		
TORS	60	72	61	74	0	180		
TORS	62	76	78	80	0	180		
TORS	82	94	83	88	0	180		
TORS	85	89	86	90	0	180		
TORS	90	91	94	95	0	180		
EXNB	0	0	0	0	0	0	0	0
CONV	2	0	0	0	0.1	0	0	0
MINI	9	0	500	0	0	0	0	0

Command file for the JBW-stochastic dynamics

```
ligandJBW.dat
ligandJBW.out
MMOD      0     1
SOLV      3     1     0     0     0     0     0     0
EXNB      0     0     0     0     0     0     0     0
FFLD      3     1     0     0     1     0     0     0
MCNV      2    25
MCSD      1     0     0     1   300
READ
#JBW/SD-body
ITOR     12     1     7
ITOR      8     2    31
ITOR      3     4    22
ITOR      1     5    10
ITOR      1     7    97
ITOR      2     8    84
ITOR      5    10    15
ITOR      5    10    17
ITOR     10    15    16
ITOR     10    15    18
ITOR      4    22    23
ITOR     22    23    25
ITOR     78    26    27
ITOR     26    27    30
ITOR      2    31    32
ITOR      2    31   110
ITOR     95    33    56
ITOR     33    34    39
ITOR     34    35    48
ITOR     36    37    43
ITOR     34    39    49
ITOR     34    39    51
ITOR     37    43    46
ITOR     43    46    47
ITOR     35    48    99
ITOR     39    49    50
ITOR     39    49    52
ITOR     33    56    57
ITOR     58    57    56
ITOR     58    59    68
ITOR     59    60    72
ITOR     60    61    74
ITOR     57    62    76
ITOR     60    72    73
ITOR     61    74    75
ITOR     62    76    77
ITOR     22    78    26
ITOR     22    78    80
ITOR     78    80    81
ITOR     83    82    94
ITOR     82    83    88
ITOR     83    84     8
ITOR     84    85    89
ITOR     85    86    90
ITOR     83    88    92
ITOR     85    89    98
ITOR     86    90    91
ITOR     90    91    93
ITOR     82    94    95
ITOR     33    95    94
#total nr of var. bond angles = 50
ITOR     97     7     1     5     0   180
ITOR     84     8     2    31     0   180
ITOR      2     8    84    85     0   180
ITOR     15    10     5     4     0   180
ITOR     78    22     4     5     0   180
ITOR     78    22    23    25     0   180
ITOR     23    22    78    80     0   180
ITOR     27    26    78    80     0   180
ITOR     78    26    27    30     0   180
ITOR    110    31     2     8     0   180
ITOR     49    39    34    35     0   180
ITOR     46    43    37    95     0   180
ITOR     37    43    46    47     0   180
ITOR     99    48    35    36     0   180
ITOR     57    56    33    34     0   180
ITOR     33    56    57    62     0   180
ITOR     73    72    60    61     0   180
ITOR     75    74    61    62     0   180
ITOR     77    76    62    61     0   180
ITOR     26    78    80    81     0   180
ITOR     92    88    83    84     0   180
ITOR     98    89    85    86     0   180
ITOR     91    90    86    87     0   180
ITOR     86    90    91    93     0   180
ITOR     95    94    82    87     0   180
ITOR     82    94    95    37     0   180
#total nr of var. bond torsion = 26
#JBW/SD-tail
IMPS      1     0     0     0     3
IMPO      1     1     0     0    20    60  0.05
MINI      9     0   100     0     0     0     0     0
MDIT      0     0     0     0   300     0     0     0
MDSA   5000    50    50    50     0     0     0     0
MDYN      1     0     1     0     1  9999   300     0
```

Appendix 3: Parameters added to the GLYCAM parameter file [344] to perform simulations with

"modified" carbohydrates.

ADDITIONS TO THE GLYCAM PARAMETER SET

```
BOND
CA-OS    364.00      1.380       from gaff
CA-N     398.00      1.400       from gaff
OG-CA    364.00      1.38        from gaff
N*-CA    440.20      1.37        from gaff
N*-NB    513.10      1.36        from gaff
CA-CV    473.70      1.39        from gaff
NB-NB    814.40      1.23        from gaff
CA-H4    367.00      1.08        from parm99
CV-H4    367.00      1.08        from parm99
CM-NA    440.20      1.37        same as ca-na
CM-HA    367.00      1.08        same as CA-HA

ANGL
CT-OG-CT    60.000    114.700    from glycam
OG-CT-C     60.000    109.500    same as os-ct-c (gaff)
OS-C -CA    67.662    110.765    from gaff
CA-NC-CA    70.000    112.200    same as ca-n2-ca (gaff)
CA-CA-NC    65.400    119.570    same as ca-ca-n2 (gaff)
C -CA-NC    64.993    120.000    Calculated with empirical approach
(gaff)
CA-CA-N     66.100    120.000    from gaff
CA-N -H     48.200    115.140    from gaff
CA-N -C     63.400    124.520    from gaff
N -CT-H2    50.000    109.500    same as hc-c3-n (gaff)
N -CT-OS    70.093    112.195    Calculated with empirical approach
(gaff)
AC-OG-CA    61.741    117.870    gaff
OG-CA-CA    66.100    122.030    gaff
CA-CT-C     64.141    110.965    gaff
CT-N*-CA    62.700    123.210    gaff
CT-N*-NB    65.400    121.520    gaff
N*-CA-CV    68.700    111.200    gaff
N*-NB-NB    75.400    112.230    gaff
CA-N*-NB    66.500    121.230    gaff
CA-CH-H4    50.000    120.000    gaff
CA-CV-NB    65.400    119.570    gaff
CV-NB-NB    73.900    113.530    gaff
N*-CA-H4    35.000    120.000    from Bio
H4-CV-NB    35.000    120.000    from Bio
CA-CV-H4    35.000    125.000    from Bio
CV-CA-H4    35.000    125.000    from Bio
N -C -CM    68.000    112.680    same as ca-c -n (gaff)
C -CM-NA    65.774    116.660    Calculated with empirical approach

DIHE
CA-CA-OS-CT   1          0.90          180.00          2.   from gaff
CT-N*-CA-H4   1          0.30          180.00          2.   from gaff
CT-N*-CA-CV   1          0.30          180.00          2.   from gaff
```

```
CT-N*-NB-NB     1          1.70        180.00         2.   from gaff
H4-CA-N*-NB     1          0.30        180.00         2.   from gaff
H4-CA-CV-H4     1          3.63        180.00         2.   from gaff
H4-CA-CV-NB     1          3.63        180.00         2.   from gaff
CV-CA-N*-NB     1          0.30        180.00         2    from gaff
CA-N*-NB-NB     1          1.70        180.00         2.   from gaff
N*-CA-CV-H4     1          3.625       180.00         2.   from gaff
N*-CA-CV-NB     1          3.625       180.00         2.   from gaff
N*-NB-NB-CV     1          8.00        180.00         2.   from gaff
C -CM-NA-C      1          0.30        180.00         2.   from gaff
C -CM-NA-H      1          0.30        180.00         2.   from gaff
C -NA-CM-CM     1          0.30        180.00         2.   from gaff
CM-CM-NA-H      1          0.30        180.00         2.   from gaff
```

Appendix 4: getrst.in

```
Constant Volume Minimization
 # Control section
 &cntrl
  ntpr = 1,
  ntwr = 1,
  ntb = 1,
  maxcyc = 1 ntr = 0,
  imin = 1,
 &end
END
END
```

Appendix 5: minres.in

Constant Volume Minimization
 # Control section
 &cntrl
  ntpr = 100,
  ntb = 1,
  maxcyc = 400,
  ntr = 1,
  imin = 1,
  ntf = 2,
  ntc = 2,
  ncyc = 200
 &end
Group input for restrain
 5.0
 RES 1
END
END

Appendix 6: annealingNEW.in

```
group info
 5.0
 RES 1 1
END
END
 &cntrl
   imin = 0,
   nmropt = 1,
   ntpr = 500,
   ntwx = 500,
   ntf = 2,
   ntc = 2,
   ntb = 1,
   igb = 0,
   ntr =1,
   nstlim = 75000,
   dt = 0.002,
   ntt = 1,
   tautp = 1.0
   temp0 = 300.0, tempi = 100.0,
 &end
 &wt
  type = 'TEMP0', istep1 = 0, istep2 = 25000,
            value1 = 100.0, value2 = 300.0,
 &end
 &wt
 type = 'TEMP0', istep1 = 25001, istep2 = 50000,
            value1 = 300.0, value2 = 300.0,
 &end
 &wt
 type = 'TEMP0', istep1 = 50001, istep2 = 75000,
            value1 = 300.0, value2 = 100.0,
 &end
 &wt
 type = 'END',
 &end
```

Appendix 7: minAN.in

```
&cntrl
 imin=1,
 ntc = 2, ntf = 2,
 ntb = 1,
 igb = 0,
 cut = 8, nsnb = 999999, ntmin = 1,
 maxcyc = 500, ncyc = 300,
 ntpr = 100,
 ntwx = 100,
 ntwr = 100
&end
```

Appendix 8: ANallinone.in

```
&cntrl
  imin = 0,
  nmropt = 1,
  ntpr = 500,
  ntwx = 500,
  ntf = 2,
  ntc = 2,
  ntb = 1,
  igb = 0,
  ntr =0,
  nstlim = 1300000,
  dt = 0.002,
  ntt = 1,
  tautp = 1.0
  temp0 = 400.0,
  tempi = 100.0,
&end

&wt
 type = 'TEMP0', istep1 = 0, istep2 = 25000,
```

```
                value1 = 100.0, value2 = 400.0,
&end
&wt
 type = 'TEMP0', istep1 = 25001, istep2 = 125000,
                value1 = 400.0, value2 = 400.0,
&end
&wt
 type = 'TEMP0', istep1 = 125001, istep2 = 150000,
                value1 = 400.0, value2 = 300.0,
&end
&wt
 type = 'TEMP0', istep1 = 150001, istep2 = 650000,
                value1 = 300.0, value2 = 300.0,
&end
&wt
 type = 'TEMP0', istep1 = 650001, istep2 = 675000,
                value1 = 300.0, value2 = 400.0,
&end
&wt
 type = 'TEMP0', istep1 = 675001, istep2 = 775000,
                value1 = 400.0, value2 = 400.0,
&end
&wt
 type = 'TEMP0', istep1 = 775001, istep2 = 800000,
                value1 = 400.0, value2 = 300.0,
&end
&wt
 type = 'TEMP0', istep1 = 800001, istep2 = 1300000,
                value1 = 300.0, value2 = 300.0,
&end
&wt
 type = 'END',
&end
```

Appendix 9: Three-letter code and affinity towards the target of the ligands used for the development of the QSAR models are listed below. Due to patenting reasons, the structures cannot be shown here, but can be obtained under written request to the Institute of Molecular Pharmacy of the University of Basel.

ligand	delta G exp (kcal/mol)
A04	-3.85
A05	-4.00
A06	-4.04
A07	-4.63
A08	-4.80
A09	-3.58
A10	-4.07
A11	-2.28
A12	-2.85
A13	-3.87
A14	-3.91
A15	-3.92
A16	-3.61
A17	-3.87
A18	-3.28
A19	-2.28
A20	-3.48
A21	-4.74
A22	-3.49
A23	-3.73
A24	-2.28

A25	-5.15
A26	-3.81
A27	-5.28
A28	-3.32
A29	-4.42
A30	-4.32
A31	-3.71
A32	-4.78
A33	-2.28
A34	-2.28
A35	-2.28
A36	-4.07
A37	-3.45
A38	-5.05
A39	-4.74
A40	-3.99
A41	-4.58
A42	-5.09
A43	-4.13
A44	-5.25
A45	-4.89
A46	-4.00
A47	-4.73
A48	-3.58
A49	-4.58
A50	-4.68
A51	-5.30

A52	-4.58
A53	-5.71
A54	-4.73
A55	-2.28
A56	-2.28
A57	-3.30
A58	-3.42
A59	-4.16
A60	-4.26
A61	-2.28
A62	-4.31
A63	-2.28
A64	-4.33
A65	-4.30
A66	-2.28
A67	-4.84
A68	-5.23
A69	-4.47
A70	-5.19
A71	-4.41
A72	-4.75
A73	-4.09
A74	-4.76
A75	-5.91
A76	-5.07
A77	-5.10
A78	-4.65

A79	-5.05
A80	-4.35
A81	-4.23
A82	-3.89
A83	-5.06
A84	-4.18
A85	-4.25
A86	-4.95
A87	-5.48
A88	-2.28
A89	-3.84
A90	-2.28
A91	-4.96
A92	-4.55
A93	-4.83
A94	-2.28
A95	-3.88
A96	-4.80
A97	-5.99
A98	-6.19
A99	-5.48
B01	-5.40
B02	-4.97
B03	-6.12
B04	-5.73
B05	-5.92
B06	-5.34

B07	-5.43
B08	-3.95
B09	-5.84
B10	-4.71
B11	-2.28
B12	-2.28
B13	-6.02
B14	-4.60
B15	-3.80
B16	-4.48
B17	-5.35
B18	-4.34
B19	-2.28
B20	-3.98
B21	-3.02
B22	-4.23
B23	-3.69
B24	-5.12
B25	-3.94
B26	-2.28
B27	-2.28
B28	-4.23
B29	-4.40
B30	-5.99
B31	-6.04
B32	-6.04
B33	-3.95

B34	-4.35
B35	-6.43
B36	-3.96
B37	-4.11
B38	-4.76
B39	-4.00
B40	-4.22
B41	-6.08
B42	-4.75
B43	-5.54
B44	-4.05
B45	-5.61
B46	-2.28
B47	-3.26
B48	-5.27
B49	-2.28
B50	-3.05
B51	-3.24
B52	-2.28
B53	-5.82
B54	-5.97
B55	-5.18
B56	-3.30
B57	-5.49
B58	-4.03
B59	-5.51
B60	-3.57

B61	-3.88
B62	-5.70
B63	-3.70
B64	-6.27
B65	-5.92
B66	-5.73
B67	-6.02
B68	-5.91
B69	-3.61
B70	-2.28
B71	-6.17
B72	-4.59
B73	-5.40
B74	-3.70
B75	-5.94
B76	-5.50
B77	-4.72
B78	-5.68
B79	-6.02
B80	-4.38
B81	-5.72
B82	-6.76
B83	-2.28
B84	-2.28
B85	-5.32
B86	-2.28
B87	-4.33

B88	-3.62
B89	-3.29
B90	-3.92
B91	-4.69
B92	-2.28
B93	-3.51
B94	-4.95
B95	-3.32
B96	-2.28
B97	-5.19
B98	-4.60
B99	-2.28
C01	-2.28
C02	-3.06
C03	-2.28
C04	-5.42
C05	-5.18
C06	-3.41
C07	-2.95
C08	-3.10
C09	-2.94
C10	-2.28
C11	-3.56
C12	-3.27
C13	-3.89
C14	-6.19
C15	-2.28

C16	-5.03
C17	-4.78
C18	-3.99
C19	-3.70
C20	-3.81
C21	-5.12
C22	-4.11
C23	-6.64
C24	-4.97
C25	-6.70
C26	-6.46
C27	-4.96
C28	-5.42
C29	-5.85
C30	-2.85
C31	-4.99
C32	-2.85
C33	-6.46
C34	-6.64
C35	-6.19
C36	-5.72
C37	-5.91
C38	-4.71

Appendix 10: List of the new torsions defined in the parameter file of *Yeti* [248,352].

Atom types				Barrier height [kcal/mol]	Periodicity	Phase shift [°]
CA	*CA*	*N*	*ON*	*12.0*	*2*	*180.0*
CT	*OX*	*C*	*CT*	*6.0*	*2*	*180.0*
C	CT	SH	HS	1.5	3	0.0
CT	OX	C	N	6.0	2	180.0
HC	CT	OX	C	0.0	3	0.0
HC	C	C	CT	12.0	2	180.0
CA	C	N	CT	20.0	2	180.0
OE	CT	C	O2	0.0	3	180.0
O	C	N	CT	20.0	2	180.0
CT	N	SS	CT	0.0	3	0.0
CT	CT	SS	N	0.0	3	0.0
CA	CT	OH	HO	1.0	3	0.0
C	N*	CT	CT	0.0	6	0.0
C	N*	C	CB	5.8	4	180.0
CA	CB	C	O	12.0	4	180.0
HC	CT	OH	HO	1.0	3	0.0
CT	OE	SS	O	0.0	3	0.0
CT	CT	OE	SS	0.0	3	0.0
CT	OX	C	CA	6.0	2	180.0
CT	OE	CT	C	4.0	3	0.0
CT	OE	CT	CT	4.0	3	0.0

Appendix 11: constraints used during the minimization step in the docking protocol (cf. *Chapter* 3.2.5.1).

Atoms / Group	Distance [Å]	Parabolic force constant [kcal/(mol.Å^2)]
Gal-6OH / Glu92	2.0	20.0
Fuc-3OH / Ca^{2+}	2.5	20.0
COOH / Arg97	2.2	20.0
Gal-4OH / Tyr94	2.0	20.0

Appendix 12: settings adopted during the MC-protocol (cf. *Chapter* 3.2.5.1).

Metal center settings	comment
octahedron	geometric property for metal center
LFSE = 0.0	ligand field stabilization: not used
deltaH = 0.0	not used
w(r) = 1.0	symmetry neglected, distance dependent selection
select alternate ligand set	Fuc–4OH selected
Minimizer settings	**comment**
Center of rotation = Fuc-4OH	Rotational center for MC trials
Weighting	**comment**
HB = 1.2, Met = 1.2	energy weights for E[lig*] contributions in the selection process

Appendix 13: Listing of the 12 ligands carrying CM-1 charges [326,327] used to test the influence of the solvent on the docking mode.

A02	A06	A22	A68	B17	B37
A05	A16	A57	A95	B34	B44

Appendix 14: minresCPL.in

Constant Volume Minimization
 # Control section
 &cntrl
  ntpr = 100,
  ntb = 1,
  maxcyc = 500,
  ntr = 1,
  imin = 1,
  ntf = 2,
  ntc = 2,
  ncyc = 10
 &end
Group input for restrain
 5.00
 RES 1 159
END
END

Appendix 15: md.in


```
&cntrl
  imin = 0,
  nmropt = 1,
  ntpr = 500,
  ntwx = 500,
  ntf = 2,
  ntc = 2,
  ntb = 1,
  igb = 0,
  ntr =0,
  nstlim = 750000,
  dt = 0.002,
  ntt = 1,
  tautp = 4.0
  temp0 = 300.0,
  tempi = 100.0,
  lastist = 8000000, lastrst = 8000000,
&end

 &wt
 type = 'TEMP0', istep1 = 0, istep2 = 50000,
          value1 = 100.0, value2 = 300.0,
 &end
 &wt
 type = 'TEMP0', istep1 = 50001, istep2 = 750000,
          value1 = 300.0, value2 = 300.0,
 &end
 &wt
 type = 'END',
 &end
```

Appendix 16: Listing of the 16 ligands used for testing the influence of the two partial-charge models on the docking mode.

A04	A13	A16	A25	A44	A67	A95	B17
A09	A14	A20	A29	A45	A91	B07	B22

Appendix 17: Ligands used to study the reverse docking mode.

3	**36**	**66**	**67**	**68g**	**68j**	**90**	B45
B68	B82						

Appendix 18: Ligands submitted to a MD simulation in the presence of the protein.

3	**26**	**36**	**66**	**67**	**74**	A41	A57
B02	B31	B45	B57	B68	B71		

Appendix 19: Desolvation energies calculated by AMSOL [326] for ligand carrying CM-1 charges

[327] and by the module *Prepare* of *Quasar* [270–272,351] for the ligand carrying ESP-MNDO

[331].


```
Ligand  dG(slv) CM-1  dG(slv) ESP-MNDO
-------------------------------------------------------

* A61 :   -72.905        -72.670
* A81 :   -75.739        -76.792
* A15 :   -72.423        -73.552
* A06 :   -79.438        -81.070
* A10 :   -72.347        -76.058
* A23 :   -70.169        -73.031
* C06 :   -94.893        -90.114
* B98 :   -92.699        -90.427
* B93 :   -82.528        -81.758
* B91 :   -85.612        -81.696
* C31 :   -59.627        -63.514
* C02 :   -89.867        -88.626
* C03 :   -93.386        -91.609
* A60 :   -83.492        -83.695
* A93 :   -65.991        -73.902
* C28 :   -72.793        -74.116
* C36 :   -70.363        -75.508
* A13 :   -67.844        -70.974
* C22 :   -65.353        -66.777
* A38 :   -65.792        -67.466
* C32 :   -61.335        -67.109
* C12 :   -89.390        -84.623
* B17 :   -79.202        -84.178
* C26 :   -65.215        -71.227
* C30 :   -68.074        -76.235
* B43 :   -73.555        -75.843
* C25 :   -71.105        -73.743
* B56 :   -73.468        -72.239
* B06 :  -106.994        -94.304
* B38 :   -87.248        -87.497
* A51 :   -90.486        -91.108
* A67 :   -87.052        -89.922
* A70 :  -100.179        -93.814
* B24 :   -95.451        -93.244
* A39 :   -88.717        -87.380
* A08 :   -89.243        -89.766
* A04 :   -87.953        -87.899
* B23 :  -102.763        -96.509
* B60 :   -94.386        -92.603
* C10 :   -89.424        -88.998
* B88 :   -89.448        -87.151
* A05 :   -86.406        -83.964
* B89 :   -93.170        -88.992
* B19 :   -87.940        -90.487
```

```
* B86 :    -87.949          -86.037
* B04 :    -71.917          -72.074
* B45 :    -74.882          -72.924
* C14 :    -70.001          -72.438
* B81 :    -72.108          -73.168
* B65 :    -71.561          -72.727
* C21 :    -72.562          -72.844
* B85 :    -72.413          -73.578
* A47 :    -68.133          -67.670
* A78 :    -73.995          -74.541
* B77 :    -68.945          -72.951
  C34 :    -65.416          -78.895
  B34 :    -70.843          -70.423
  A16 :    -69.434          -71.609
  A07 :    -66.735          -70.651
  B37 :    -69.307          -67.568
  B75 :    -65.866          -69.011
  B47 :    -73.634          -73.500
  C23 :    -82.422          -81.141
  A17 :    -70.807          -72.181
  A37 :    -90.249          -89.124
  A64 :    -83.219          -82.906
  A72 :    -85.487          -82.963
  A98 :    -74.465          -73.521
  B30 :    -69.367          -71.742
  B33 :    -62.143          -65.591
  B54 :    -72.100          -71.616
  B64 :    -71.072          -72.677
  B80 :    -69.166          -76.275
  B92 :    -83.897          -82.098
  B94 :    -88.168          -81.151
  B97 :    -98.119          -91.664
  C09 :    -89.461          -86.419
  C11 :    -86.921          -82.228
  C17 :    -63.756          -66.010
  C19 :    -68.557          -75.824
  C33 :    -68.247          -75.572
  A79 :    -77.659          -80.165
  A63 :    -72.421          -72.852
  A73 :    -68.870          -70.590
  A97 :    -71.260          -73.966
  A80 :    -77.243          -76.962
  A95 :    -77.345          -80.739
  B36 :    -75.984          -76.050
  B42 :    -86.510          -80.476
  B44 :    -75.231          -75.778
  B55 :    -78.118          -77.310
  B62 :    -77.488          -76.592
  C01 :    -88.315          -88.016
  B99 :    -91.995          -89.763
  C07 :    -90.355          -86.738
  B25 :    -89.573          -85.250
  A96 :    -64.976          -71.766
  C24 :    -67.742          -74.341
  C37 :    -65.475          -70.404
  B26 :    -59.372          -64.925
```

```
A19 :   -62.194        -68.650
A94 :   -66.646        -71.188
A11 :   -61.091        -67.718
A57 :   -72.113        -74.476
C18 :   -66.599        -72.192
A26 :   -76.694        -77.595
C04 :   -71.986        -71.418
A49 :   -71.471        -72.566
A44 :   -73.574        -79.373
A12 :   -69.771        -76.122
A88 :   -72.505        -76.973
C15 :   -62.632        -64.260
A34 :   -65.379        -70.853
B84 :   -64.985        -69.908
A24 :   -65.633        -70.149
B83 :   -67.031        -73.602
B95 :   -48.681        -78.362
A48 :   -74.092        -73.929
A14 :   -69.955        -73.985
C20 :   -64.665        -69.261
B01 :   -58.073        -63.065
A29 :   -69.142        -70.403
C38 :   -71.462        -70.171
A45 :   -70.970        -76.327
A42 :   -64.676        -67.908
C16 :   -67.355        -69.491
C13 :   -84.709        -81.067
A69 :   -83.742        -81.220
B27 :   -86.915        -78.828
B02 :   -81.317        -75.263
A27 :   -68.262        -73.130
B09 :   -69.965        -70.736
B48 :   -95.333        -94.180
B40 :   -88.742        -88.166
B07 :   -90.930        -92.590
A76 :   -89.151        -88.805
A54 :   -85.994        -87.152
B05 :  -100.683        -95.841
A87 :   -98.563        -92.077
A68 :   -91.230        -90.438
A53 :   -97.667        -93.597
B72 :   -94.228        -91.921
A52 :   -85.759        -87.235
B22 :   -90.686        -92.150
A40 :   -92.659        -87.404
B61 :   -92.437        -93.091
A20 :   -95.332        -95.552
A22 :   -93.082        -92.467
A09 :   -88.359        -87.950
B21 :   -99.788        -93.593
B51 :   -91.571        -86.405
B52 :   -92.931        -86.693
B58 :   -90.447        -89.073
B18 :   -87.879        -89.284
A59 :   -90.034        -86.686
B87 :   -87.166        -86.991
```

```
B20 :   -91.570        -87.504
B39 :   -83.208        -84.053
A58 :   -87.740        -83.548
A56 :   -88.731        -85.482
C29 :   -72.772        -71.462
B57 :   -79.383        -74.659
B53 :   -74.806        -74.729
B03 :   -68.927        -70.030
B31 :   -68.992        -71.431
B67 :   -71.139        -72.211
B73 :   -71.105        -71.955
B32 :   -71.139        -72.299
B66 :   -74.738        -74.336
A91 :   -74.584        -76.287
A74 :   -74.451        -70.655
A50 :   -73.547        -73.208
A83 :   -75.936        -73.217
A46 :   -64.546        -71.521
B63 :   -69.927        -71.708
A33 :   -67.390        -76.250
B82 :   -72.324        -79.921
B74 :   -77.145        -78.633
B13 :   -77.959        -77.918
B41 :   -87.174        -81.741
B68 :   -81.480        -83.432
----------------------------------------------

Min. :  -106.994       -96.509
Max. :   -48.681       -63.065
Rng. :    58.313        33.444
----------------------------------------------
```

Appendix 20: Test-/Training-set arrangement of the ligands in the different *Quasar-models*.

Q11

training				test	
A07	A69	B36	C07	A06	B93
A11	A72	B42	C09	A10	B98
A12	A73	B43	C11	A13	B99
A14	A79	B44	C13	A15	C02
A17	A80	B47	C15	A16	C04
A24	A88	B54	C16	A19	C06
A26	A95	B55	C17	A23	C12
A27	A96	B62	C19	A38	C18
A29	A97	B64	C20	A44	C22
A34	A98	B80	C23	A60	C28
A37	B01	B83	C24	A61	C30
A42	B02	B84	C25	A81	C31
A45	B17	B92	C26	A93	C34
A48	B25	B94	C32	A94	C36
A49	B26	B95	C33	B34	
A57	B27	B97	C37	B37	
A63	B30	C01	C38	B75	
A64	B33	C03	C06	B91	

Q12

training				test	
A11	A72	B36	C06	A06	C32
A14	A73	B37	C07	A07	C34
A16	A79	B42	C09	A10	C36
A17	A80	B43	C11	A12	
A19	A81	B47	C13	A13	
A24	A88	B54	C15	A15	
A26	A93	B55	C16	A23	
A27	A94	B62	C17	A38	
A29	A95	B64	C18	A60	
A34	A97	B80	C19	A61	
A37	A98	B83	C20	A96	
A42	B01	B84	C23	B44	
A44	B02	B92	C24	B75	
A45	B17	B94	C25	B91	
A48	B25	B95	C26	B93	
A49	B26	B97	C30	B98	
A57	B27	B99	C31	C02	
A63	B30	C01	C33	C12	
A64	B33	C03	C37	C22	
A69	B34	C04	C38	C28	

Q13

training				test	
A07	A63	B33	C04	A06	C28
A11	A64	B36	C06	A10	C30
A12	A69	B37	C07	A13	C31
A14	A72	B42	C09	A15	C32
A16	A73	B43	C11	A23	C34
A17	A79	B44	C13	A38	C36
A19	A80	B47	C15	A60	
A24	A81	B54	C16	A61	
A26	A88	B55	C17	A93	
A27	A94	B62	C18	B34	
A29	A95	B64	C19	B75	
A34	A96	B80	C20	B91	
A37	A97	B83	C23	B93	
A42	A98	B84	C24	B97	
A44	B01	B92	C25	B98	
A45	B17	B94	C26	C02	
A48	B25	B95	C33	C03	
A49	B26	B99	C37	C12	
A57	B30	C01	C38	C22	

Q14

training				test	
A07	A63	B33	B99	A06	C22
A11	A64	B34	C01	A10	C25
A12	A69	B36	C04	A13	C26
A14	A72	B37	C07	A15	C28
A16	A73	B42	C09	A23	C30
A17	A79	B44	C11	A38	C31
A19	A80	B47	C13	A60	C32
A24	A88	B54	C15	A61	C36
A26	A94	B55	C16	A81	
A27	A95	B62	C17	A93	
A29	A96	B64	C18	B17	
A34	A97	B75	C19	B43	
A37	A98	B80	C20	B91	
A42	B01	B83	C23	B93	
A44	B02	B84	C24	B98	
A45	B25	B92	C33	C02	
A48	B26	B94	C34	C03	
A49	B27	B95	C37	C06	
A57	B30	B97	C38	C12	

Q15–Q17					
training				**test**	
A07	A68	B30	B80	A04	B45
A09	A69	B31	B82	A05	B60
A11	A72	B32	B83	A06	B65
A12	A73	B33	B84	A08	B75
A14	A74	B36	B87	A10	B77
A17	A76	B39	B92	A13	B81
A20	A79	B40	B94	A15	B85
A22	A80	B41	B95	A16	B86
A24	A83	B42	B97	A19	B88
A26	A87	B43	C01	A23	B89
A27	A88	B44	C03	A38	B91
A29	A91	B47	C07	A39	B93
A33	A95	B48	C09	A44	B98
A34	A96	B51	C11	A47	B99
A37	A97	B52	C13	A51	C02
A40	A98	B53	C15	A60	C04
A42	B01	B54	C16	A61	C06
A45	B02	B55	C17	A67	C10
A46	B03	B56	C19	A70	C12
A48	B05	B57	C20	A78	C14
A49	B06	B58	C23	A81	C18
A50	B07	B61	C24	A93	C21
A52	B13	B62	C25	A94	C22
A53	B17	B63	C26	B04	C28
A54	B18	B64	C29	B09	C30
A56	B20	B66	C32	B19	C31
A57	B21	B67	C33	B23	C34
A58	B22	B68	C37	B24	C36
A59	B25	B72	C38	B34	
A63	B26	B73		B37	
A64	B27	B74		B38	

Q18–Q20					
training				**test**	
A07	A63	B27	B75	A04	B81
A09	A64	B30	B80	A05	B85
A11	A68	B31	B82	A06	B86
A12	A69	B32	B83	A08	B88
A14	A72	B33	B84	A10	B89
A16	A73	B34	B87	A13	B91
A17	A74	B36	B92	A15	B93
A19	A76	B37	B94	A23	B98
A20	A79	B39	B95	A38	C02
A22	A80	B40	B97	A39	C03
A24	A83	B41	B99	A47	C06
A26	A87	B42	C01	A51	C10
A27	A88	B44	C04	A60	C12
A29	A91	B47	C07	A61	C14
A33	A94	B48	C09	A67	C21
A34	A95	B51	C11	A70	C22
A37	A96	B52	C13	A78	C25
A40	A97	B53	C15	A81	C26
A42	A98	B54	C16	A93	C28
A44	B01	B55	C17	B04	C30
A45	B02	B57	C18	B06	C31
A46	B03	B58	C19	B17	C32
A48	B05	B61	C20	B19	C36
A49	B07	B62	C23	B23	
A50	B09	B63	C24	B24	
A52	B13	B64	C29	B38	
A53	B18	B66	C33	B43	
A54	B20	B67	C34	B45	
A56	B21	B68	C37	B56	
A57	B22	B72	C38	B60	
A58	B25	B73		B65	
A59	B26	B74		B77	

Q21–Q23					
training				test	
A07	A63	B25	B73	A04	B85
A09	A64	B26	B74	A05	B86
A11	A68	B27	B80	A06	B88
A12	A69	B30	B82	A08	B89
A14	A72	B31	B83	A10	B91
A16	A73	B32	B84	A13	B93
A17	A74	B33	B87	A15	B97
A19	A76	B36	B92	A23	B98
A20	A79	B37	B94	A38	C02
A22	A80	B39	B95	A39	C03
A24	A81	B40	B99	A47	C10
A26	A83	B41	C01	A51	C12
A27	A87	B42	C04	A60	C14
A29	A88	B43	C06	A61	C21
A33	A91	B44	C07	A67	C22
A34	A94	B47	C09	A70	C28
A37	A95	B48	C11	A78	C30
A40	A96	B51	C13	A93	C31
A42	A97	B52	C15	B04	C32
A44	A98	B53	C16	B09	C34
A45	B01	B54	C17	B19	C36
A46	B02	B55	C18	B23	
A48	B03	B57	C19	B24	
A49	B05	B58	C20	B34	
A50	B06	B61	C23	B38	
A52	B07	B62	C24	B45	
A53	B13	B63	C25	B56	
A54	B17	B64	C26	B60	
A56	B18	B66	C29	B65	
A57	B20	B67	C33	B75	
A58	B21	B68	C37	B77	
A59	B22	B72	C38	B81	

Q24–Q26					
training				test	
A07	A63	B27	B75	A04	B81
A09	A64	B30	B80	A05	B85
A11	A68	B31	B82	A06	B86
A12	A69	B32	B83	A08	B88
A14	A72	B33	B84	A10	B89
A16	A73	B34	B87	A13	B91
A17	A74	B36	B92	A15	B93
A19	A76	B37	B94	A23	B98
A20	A79	B39	B95	A38	C02
A22	A80	B40	B97	A39	C03
A24	A83	B41	B99	A47	C06
A26	A87	B42	C01	A51	C10
A27	A88	B44	C04	A60	C12
A29	A91	B47	C07	A61	C14
A33	A94	B48	C09	A67	C21
A34	A95	B51	C11	A70	C22
A37	A96	B52	C13	A78	C25
A40	A97	B53	C15	A81	C26
A42	A98	B54	C16	A93	C28
A44	B01	B55	C17	B04	C30
A45	B02	B57	C18	B06	C31
A46	B03	B58	C19	B17	C32
A48	B05	B61	C20	B19	C36
A49	B07	B62	C23	B23	
A50	B09	B63	C24	B24	
A52	B13	B64	C29	B38	
A53	B18	B66	C33	B43	
A54	B20	B67	C34	B45	
A56	B21	B68	C37	B56	
A57	B22	B72	C38	B60	
A58	B25	B73		B65	
A59	B26	B74		B77	

7 References

[1]   M.P. Bevilacqua, R.M. Nelson, G. Mannori, O. Cecconi, *Annu. Rev. Med.* **1994**, *45*, 361-378.
[2]   G.S. Kansas, *Blood* **1996**, *88*, 3259-3287.
[3]   R.P. McEver, K.L. Moore, R.D. Cummings, *J. Biol. Chem.* **1995**, *270*, 11025-11028.
[4]   N. Kojima, K. Handa, W. Newman, S. Hakomori, *Biochem. Biophys. Res. Commun.* **1992**, *182*, 1288-1295.
[5]   S. Hakomori, *Proc. Nat. Acad. Sci. USA,* **2002**, *99*, 10231-10233.
[6]   K. Ley., *Leukocyte recruitment as seen by intravital microscopy*, Oxford University press, Oxford, **2001**.
[7]   C.N. Serhan, P.A. Ward, *Molecular and Cellular basis of Inflammation* **1998**, Humana Press Inc.
[8]   T.J. Schall, K.B. Bacon, *Curr. Opin. Immunol.* **1994**, *6*, 865-873.
[9]   T. Springer, *Annu. Rev. Physiol.* **1995**, *57*, 827-872.
[10]   E.C. Butcher, *Cell* **1991**, *67*, 1033-1036.
[11]   M.P. Bevilacqua, M. Gimbrone, B. Seed, *Science* **1989**, *243*, 1160-1165.
[12]   G.I. Johnston, R.G. Cook, R.P. McEver, *Cell* **1989**, *56*, 1033-1044.
[13]   J.G. Geng, M.P. Bevilacqua, K.L. Moore, T.M. McIntire, S.M. Prescott, J.M. Kim, G.A. Bliss, G. Zimmerman, R.P. McEver, *Nature* **1990**, *343*, 757-760.
[14]   R. Hattori, K.K. Hamilton, R.D. Fugate, R.P. McEver, P.J. Sims, *J. Biol. Chem.* **1989**, *264*, 7768-7771.
[15]   M.P. Bevilacqua, J.S. Pober, D.L. Mendrick, R.S. Cotran, M.A. Gimbrone Jr., *Proc. Natl. Acad. Sci. USA* **1987**, *84*, 9238-9242.
[16]   J.S. Pober, L.A. Lapierre, A.H. Stolpen, T.A. Brock, T. Springer, W. Fiers, M.P. Bevilacqua, D.L. Mendrick, M.A. Gimbrone Jr., *J. Immunol.* **1987**, *138*, 3319-3324.
[17]   M.P. Bevilacqua, M. Gimbrone, B. Seed, *Science* **1989**, *243*, 1160-1165.
[18]   S. Nakamori, H. Furukawa, M. Hiratsuka, T. Iwanaga, S. Imaoka, O. Ishikawa, T. Kabuto, Y. Sasaki, M. Kameyama, S. Ishiguro, T. Irimura, *J Clin Oncol.* **1997**, *15*, 816-825.
[19]   A. Atherton, G.V.R. Born, *J. Physiol.* **1972**, *222*, 447-474.
[20]   K. Ley, P. Gaehtgens, C. Fennie, M.S. Singer, L.A. Lasky, S.D. Rosen, *Blood* **1991**, *77*, 2553.
[21]   L.J. Picker, R.A. Warnock, A.R. Burns, C.M. Doerschuk, E.L. Berg, E.C. Butcher, *Cell* **1991**, *66*, 921-933.
[22]   B. Walcheck, K.L. Moore, R.P. McEver, K. Kishimoto, *J. Clin. Invest.* **1996**, *98*, 1081-1087.
[23]   R.F. Bargatze, S. Kurk, E.C. Butcher, M.A. Jutila, *J. Exp. Med.* **1994**, *180*, 1785-1792.
[24]   U.H. von Adrian, J.D. Chambers, L.M. McEvoy, R.F. Bargatze, K.E. Arfors, E.C. Butcher, *Proc. Natl. Acad. Sci. USA* **1991**, *88*, 7538.
[25]   M.B. Lawrence, T. Springer, *Cell* **1991**, *65*, 859-873.
[26]   X.L. Ma, P.S. Tsao, A.M. Lefer, *J. Clin. Invest.* **1991**, *88*, 1237-1243.
[27]   L.A. Hernandez, M.B. Grisham, K.E. Twohig, J.M. Arfors, J.M. Harlan, D.N. Granger, *Am. J. Physiol.* **1987**, *253*, H699-H703.

[28] A. Seekamp, G.O. Till, M.S. Mulligan, J.C. Paulson, D.C. Anderson, M. Miyasaka, P.A. Ward, *Am. J. Pathol.* **1994**, *144*, 592-598.

[29] B.S. Coller, *J. Clin. Invest.* **1997**, *99*, 1467-1471.

[30] H. Matsuno, J.M. Stassen, J. Vermylen, H. Deckmyn, *Circulation* **1994**, *90*, 2203-2206.

[31] O. Abbassi, K. Kishimoto, T.M. McIntire, D.C. Anderson, C.W. Smith, *J. Clin. Invest.* **1993**, *92*, 2719-2730.

[32] G. Kaplanski, C. Franarier, O. Tissot, A. Pierres, A.-N. Benoliel, S. Kaplanski, P. Bongrand, *Biophys. J.* **1993**, *64*, 1922-1933.

[33] M.B. Lawrence, T. Springer, *J. Immunol.* **1993**, *151*, 6338-6346.

[34] M. Dore, R.J. Korthuis, D.N. Granger, M.L. Entman, C.W. Smith, *Blood* **1993**, *82*, 1308-1316.

[35] D.A. Steeber, N.E. Green, S. Sato, T. Tedder, *J. Immunol.* **1996**, *157*, 1096-1106.

[36] M.L. Arbones, D.C. Ord, K. Ley, H. Ratesch, C. Maynard-Curry, G. Otten, D.J. Capon, T. Tedder, *Immunity* **1994**, *1*, 247-260.

[37] M.D. Catalina, M.C. Carroll, H. Arizpe, A. Takashima, P. Estess, M.H. Siegelman, *J. Exp. Med.* **1996**, *184*, 2341-2351.

[38] T. Tedder, D.A. Steeber, P. Pizcueta, *J. Exp. Med.* **1995**, *181*, 2259-2264.

[39] J.C. Xu, I.S. Grewal, G.P. Geba, R.A. Flavell, *J. Exp. Med.* **1996**, *183*, 589-598.

[40] T.N. Mayadas, R.C. Jonson, H. Rayburn, R.O. Hynes, D. Wagner, *Cell* **1993**, *74*, 541-554.

[41] D.C. Bullard, L. Quin, I. Lorenzo, W.M. Quinlin, D. Vestweber, A.L. Beaudet, *J. Clin. Invest.* **1995**, *95*, 1782-1788.

[42] T.M. McIntire, *Immunity* **1994**, *1*, 709-720.

[43] D.C. Bullard, E.J. Kunkel, H. Kubo, M.J. Hicks, C.M. Doerschuk, A.L. Beaudet, *J. Exp. Med.* **1996**, *183*, 2329-2336.

[44] E.J. Kunkel, K. Ley, *Circ. Res.* **1996**, *79*, 1196-1204.

[45] P.S. Frenette, T.N. Mayadas, H. Rayburn, R.O. Hynes, D. Wagner, *Cell* **1996**, *84*, 563-574.

[46] M. P. Bevilacqua, R. M. Nelson, G. Mannori, O. Cecconi, *Annu. Rev. Med.* **1994**, *45*, 361-378.

[47] S.T. Holgate, J.K. Shute, R. Djukanovic, A.F. Walls, M.K. Church, in *Eosinophils: Biological and Clinical Aspects* (Eds.: S. Makino, T. Fukuda), CRC Press, Boca Raton, **1992**, pp. 243-260.

[48] R.W. Groves, M.H. Allen, J.N. Barker, D.O. Haskard, D.M. Macdonald, *British J. Dermatol.* **1991**, *124*, 117-123.

[49] E.E. Eriksson, X. Xie, J. Werr, P. Thoren, L. Lindbom, *J. Exp. Med.* **2001**, *194*, 205-218.

[50] M. Blaha, J. Krejsek, V. Blaha, C. Andrys, D. Vokurkova, J. Maly, M. Blazek, M. Skooepova, *Physiol. Res.* **2004**, *53*, 273-278.

[51] Z.M. Dong, A.A. Brow, D.D. Wagner, *Circulation* **2000**, *101*, 2290-2295.

[52] K. Wenzel, R. Stahn, A. Speer, K. Denner, C. Glaser, M. Affeldt, M. Moobed, A. Scheer, G. Baumann, S.B. Felix, *Biol.Chem* **1999**, *380*, 661-667.

[53] A.M. Lefer, A.S. Weyrich, M. Buerke, *Cardiovasc. Res.* **1994**, *28*, 289-294.

[54] A.S. Weyrich, L. Ma, A.M. Lefer, D.J. Lefer, K.H. Albertine, *J. Clin. Invest.* **1993**, *91*, 2620-2629.

[55] M. Buerke, A.S. Weyrich, Z.L. Zheng, F.A. Gaeta, M.J. Forrest, A.M. Lefer, *J. Clin. Invest.* **1994**, *93*, 1140-1148.

[56] H.J. Egerer K, Feist E, Albrecht A, Rudolph PE, Dorner T, Burmester GR., *Arthritis Rheum.* **2003**, *49*, 546-548.

[57]   N.S. Bloom BJ, Alario AJ, Miller LC, Schaller JG, *Rheumatol Int.* **2002**, *22*, 175-177.
[58]   A. Ates, G. Kinikli, M. Turgay, M. Duman, *Scand J Immunol.* **2004**, *59*, 315-320.
[59]   H.J. Voskuyl AE, Zwinderman AH, Paleolog EM, van der Meer FJ, Daha MR, Breedveld FC, *Ann Rheum Dis* **2003**, *62*, 407-413.
[60]   M. Wein, B.S. Bochner, *Eur. Resp. J.* **1993**, *6*, 1239-1242.
[61]   G. Steinhoff, M. Behrend, B. Schrader, *Am. J. Pathol.* **1993**, *142*, 481-488.
[62]   C. Ferran, M. Peuchmaur, M. Desruennes, *Transplantation* **1993**, *55*, 605-609.
[63]   D.M. Briscoe, F.J. Schoen, G.E. Rice, *Transplantation* **1991**, *51*, 537-547.
[64]   C. Brockmeyer, M. Ulbrecht, D.J. Schendel, *Transplantation* **1993**, *55*, 610-615.
[65]   J. Norton, J.P. Sloane, N. al-Saffar, *J. Clin. Pathol.* **1991**, *44*, 586-591.
[66]   J. Norton, J.P. Sloane, N. al-Saffar, *Clin. Exp. Immunol.* **1992**, *87*, 231-236.
[67]   L. Borsig, R. Wong, O. Hynes Richard, M. Varki Nissi, A. Varki, *Proc. Nat. Acad. Sci. USA* **2002**, *99*, 2193-2198.
[68]   A. Etzioni, M. Frydman, S. Pollack, I. Avidor, M. L. Phillips, J. C. Paulson, R. Gershoni-Baruch, *New Eng. J. Med.* **1992**, *327*, 1789-1792.
[69]   P. S. Frenette, D. D. Wagner, *Thromb.Haemostas.* **1997**, *78*, 60-64.
[70]   D. Wong, K. Dorovini-Zis, *J. Neuropath. Exp. Neurolog.* **1996**, *55*, 225-235.
[71]   K. Fassbender, S. Kaptur, P. Becker, J. Groschl, M. Hennerici, *Clin. Immunol. Immunopath.* **1998**, *89*, 54-60.
[72]   D. J. Lefer, *Annu. Rev. Pharmacol. Toxicol.* **2000**, *40*, 283-294.
[73]   L. Borsig, R. Wong, O. Hynes Richard, M. Varki Nissi, A. Varki, *Proc. Nat. Acad. Sci. USA* **2002**, *99*, 2193-2198.
[74]   L. Borsig, R. Wong, J. Feramisco, D. R. Nadeau, N. M. Varki, A. Varki, *Proc. Nat. Acad. Sci. USA* **2001**, *98*, 3352-3357.
[75]   F. Qian, D. Hanahan, I. L. Weissman, *Proc. Nat. Acad. Sci. USA* **2001**, *98*, 3976-3981.
[76]   Y. J. Kim, L. Borsig, N. M. Varki, A. Varki, *Proc. Nat. Acad. Sci. USA* **1998**, *95*, 9325-9330.
[77]   R. Kannagi, *Glycoconj. J.* **1997**, *14*, 577-584.
[78]   A. Etzioni, M. Frydman, S. Pollack, I. Avidor, M. L. Phillips, J. C. Paulson, R. Gershoni-Baruch, *New Eng. J. Med.* **1992**, *327*, 1789-1792.
[79]   R.U. Lemieux, J. Le Pendu, J.P. Carton, R. Oriol, *Am. J. Hum. Genet.* **1985**, *37*, 749-760.
[80]   D.C. Anderson, T. Springer, *Annu. Rev. Med.* **1987**, *38*, 175-192.
[81]   S.H. Barondes, D.N.W. Cooper, M.A. Gitt, H. Leffler, *J. Biol. Chem.* **1994**, *269*, 20807-20810.
[82]   M. Spiess, *Biochemistry* **1990**, *29*, 10009-10018.
[83]   W.I. Weis, R. Kahn, R. Fourme, K. Drickamer, W. Hendrickson, *Science* **1991**, *254*, 1608-1615.
[84]   W.I. Weis, K. Drickamer, W.A. Hendrickson, *Nature* **1992**, *360*, 127-134.
[85]   M.P. Bevilacqua, E.C. Butcher, B. Furie, B.C. Furie, W.M. Gallatin, M.A. Gimbrone, J.M. Harlan, T.K. Kishimoto, L.A. Lasky, R.P. McEver, J.C. Paulson, S.D. Rosen, B. Seed, M.H. Siegelman, T.A. Springer, L.M. Stoolman, T.F. Tedder, A. Varki, D.D. Wagner, I.L. Weismann, G.A. Zimmerman, *Cell* **1991**, *67*, 233.
[86]   D. V. Erbe, B. A. Wolitzky, L. G. Presta, C. R. Norton, R. J. Ramos, D. K. Burns, J. M. Rumberger, B. N. N. Rao, C. Foxall, *et al.*, *J. Cell Biol.* **1992**, *119*, 215-227.
[87]   C. Laudanna, G. Constantin, P. Baron, E. Scarpini, G. Scarlano, G. Caprini, C. Dechecchi, F. Rossi, M.A. Cassatella, G. Berton, *J. Biol. Chem.* **1994**, *269*, 4021-4026.

[88]   T.P. Patel, M.U. Nollert, R.P. McEver, *J. Cell Biol.* **1995**, *313*, 1893-1902.
[89]   S.H. Li, D.K. Burns, J.M. Rumberger, D.H. Presky, V.L. Wilkinson, M. Anostario, B.A. Wolitzky, C.R. Norton, P.C. Familletti, K.J. Kim, A.L. Goldstein, D.S. Cox, K.S. Huang, *J. Biol. Chem.* **1994**, *269*, 4431-4437.
[90]   P. Hensley, P.J. McDevitt, I. Brooks, J.J. Trill, J.A. Feild, D.E. McNulty, J.R. Connor, D.E. Griswold, N.V. Kumar, K.D. Kopple, S.A. Carr, B.J. Dalton, K. Johanson, *J. Biol. Chem.* **1994**, *269*, 23949-23958.
[91]   F. Kolbinger, J.T. Patton, G. Geisenhof, A. Aenis, X. Li, A. Katopodis, *Biochemistry* **1996**, *35*, 6385-6392.
[92]   D.V. Erbe, B.A. Wolitzky, L.G. Presta, C.R. Norton, R.J. Ramos, D.K. Burns, J.M. Rumberger, B.N.N. Rao, C. Foxall, *et al.*, *J. Cell Biol.* **1992**, *119*, 215-227.
[93]   R. Piggot, L.A. Needham, R.M. Edwards, C. Walker, C. Power, *J. Immunol.* **1991**, *147*, 130-135.
[94]   C.E. Green, D.N. Pearson, R.T. Camphausen, D.E. Staunton, S.I. Simon, *J. Immunol.* **2004**, *172*, 7780-7790.
[95]   G.S. Kansas, K.B. Saunders, K. Ley, A. Zarkzewich, R.M. Gibson, B.C. Furie, T.F. Tedder, *J.Cell.Biol.* **1994**, *124*, 609-618.
[96]   D.V. Erbe, S.R. Watson, L.G. Presta, B.A. Wolitzky, C. Foxall, B.K. Brandley, L.A. Lasky, *J. Cell Biol.* **1993**, *120:*, 1227-1235.
[97]   D.V. Erbe, B.A. Wolitzky, L.G. Presta, C.R. Norton, R.J. Ramor, D. Burns, J.M. Rumberger, R. Narasinga, C. Foxall, B.K. Brandley, L.A. Lasky, *J.Cell Biol.* **1992**, *119*, 215-227.
[98]   P.K. Mehta P, Laue TM, Erickson HP, McEver RP, *Blood* **1997**, *90*, 2381-2389.
[99]   W.S. Somers, J. Tang, G.D. Shaw, R.T. Camphausen, *Cell* **2000**, *103*, 467-479.
[100]   K. Drickamer, *J. Biol. Chem.* **1988**, *263*, 9557-9560.
[101]   E.L. Berg, M.K. Robinson, O. Mansson, E.C. Butcher, J. Magnani, *J. Biol. Chem.* **1991**, *266*, 14869-14872.
[102]   G. Walz, A. Aruffo, W. Kolanus, M.P. Bevilacqua, B. Seed, *Science* **1990** *250*, 1132.
[103]   L. Phillips, E. Nudelman, F.A. Gaeta, M. Perez, A.K. Singhal, S. Hakomori, J. Paulson, *Science* **1990**, *250*, 1130-1132.
[104]   P.R. Streeter, E.L. Berg, B.T.N. Rouse, R.F. Bargatze, E.C. Butcher, *Nature* **1988**, *331*, 41-46.
[105]   M. Nakache, E.L. Berg, P.R. Streeter, E.C. Butcher, *Nature* **1989**, *337*, 179-181.
[106]   S. Hemmerich, E.C. Butcher, S.D. Rosen, *J. Exp. Med.* **1994**, *180*, 2219-2226.
[107]   L.A. Lasky, S.D. Rosen, C. Fennie, M.S. Singer, Y. Imai, *J. Cell Biol.* **1991**, *113*, 1213-1221.
[108]   S. Baumhueter, M.S. Singer, W. Henzel, S. Hemmerich, M. Renz, S.D. Rosen, L.A. Lasky, *Science* **1993**, *262*, 436-438.
[109]   C. Sassetti, K. Tangemann, M.S. Singer, D.B. Kershaw, S.D. Rosen, *J. Exp. Med.* **1998**, *187*, 1965-1975.
[110]   M. Brustein, G. Kraal, R. Mebius, S. Watson, *J. Exp. Med.* **1992**, *176*, 1415-1419.
[111]   S.D. Rosen, A. Kikuta, *Blood* **1994**, *84*, 3766-3775.
[112]   S. Hemmerich, C.R. Bertozzi, H. Leffler, S.D. Rosen, *Biochemistry* **1994**, *33*, 4820-4829.
[113]   Y. Imai, L.A. Lasky, S.D. Rosen, *Nature* **1993**, *361*, 555-557.
[114]   K.L. Moore, N.L. Stultz, S. Diaz, D.L. Smith, R.D. Cummings, A. Varki, R.P. McEver, *J. Cell Biol.* **1992**, *118*, 445-456.
[115]   T. Pouyani, B. Seed, *Cell* **1995**, *83*, 333-343.
[116]   D. Sako, K.M. Comess, K.M. Barone, R.T. Camphausen, D.A. Cumming, G. Shaw, *Cell* **1995**, *83*, 323-331.

[117] P.P. Wilkins, K.L. Moore, R.P. McEver, S.D. Rosen, *J. Biol. Chem.* **1995**, *270*, 22677-22680.

[118] M. Schapira, *J. Cell Biol.* **1996**, *135*, 523-531.

[119] A. Levinovitz, J. Mühlhoff, S. Isenmann, D. Vestweber, *J. Cell Biol.* **1993**, *121*, 449-459.

[120] M. Lenter, A. Levinovitz, S. Isenmann, D. Vestweber, *J. Cell Biol.* **1994**, *125*, 471-481.

[121] R.R. Lobb, T.P. Patel, S.E. Goelz, R.B. Parekh, *Biochemistry* **1994**, *33*, 14815-14824.

[122] R. Lobb, G. Chi-Rosso, D.R. Leone, M.D. Rosa, S. Bixler, B.M. Newman, S. Luhowskyj, C.D. Benjamin, I.G. Dougas, S.E. Goelz, C. Hession, E.P. Chow, *J. Immunol.* **1991**, *147*, 124-129.

[123] F. Li, P.P. Wilkins, S. Crawley, J. Weinstein, D.A. Cumming, R.P. McEver, *J. Biol. Chem.* **1996**, *271*, 3255-3264.

[124] K.L. Moore, S.F. Eaton, D.E. Lyons, H.S. Liechenstein, D.A. Cumming, R.P. McEver, *J. Biol. Chem.* **1994**, *269*, 23318-23327.

[125] T.P. Patel, K.L. Moore, M.U. Nollert, R.P. McEver, *J. Clin. Invest.* **1995**, *96*, 1887-1896.

[126] D.J. Goetz, D.M. Greif, H. Ding, R.T. Camphausen, S. Howes, K.M. Comess, K.R. Snapp, G.S. Kansas, F.W. Luscinskas, *J. Cell Biol.* **1997**, *137*, 509-519.

[127] D. Asa, L. Raycroft, L. Ma, P.A. Aeed, P.S. Kaytes, A.P. Elhammer, J.G. Geng, *J. Biol. Chem.* **1995**, *270*, 11662-11670.

[128] K. Kishimoto, R.A. Warnock, M.A. Jutila, E.C. Butcher, C. Lane, D.C. Anderson, C.W. Smith, *Blood* **1991**, *78*, 805-811.

[129] D. Vestweber, J.E. Blanks, *Physiol. Rev.* **1999**, *79*, 181-213.

[130] D. Hammer, *Proc. R. Soc. Lond. B Biol. Sci.* **1988**, *234*, 55-83.

[131] D. Hammer, *Biophys. J.* **1992**, *63*, 35-57.

[132] E.C. Butcher, D.M. Lewinsohn, R.F. Bargatze, *J. Immunol.* **1987**, *138*, 4313-4321.

[133] W.J. Sanders, T.R. Katsumoto, C.R. Bertozzi, S.D. Rosen, L.L. Kiessling, *Biochemistry* **1996**, *35*, 14862-14867.

[134] C. Foxall, S. Watson, D. Dowbenko, C. Fennie, L.A. Lasky, M. Kiso, A. Hasegawa, D. Asa, B.K. Brandley, *J. Cell Biol.* **1992**, *117*, 895-902.

[135] G.S. Jacob, C. Kirmaier, S.Z. Abba, S.C. Howard, C.N. Steininger, J.K. Welply, P. Scudder, *Biochemistry* **1995**, *34*, 1210-1217.

[136] M.P. Bevilacqua, *J. Clin. Invest.* **1993**, *91*, 1157-1166.

[137] M.K. Wild, M.-C. Huang, U. Schulze-Horsel, P.A. van der Meerwe, D. Vestweber, *J. Biol. Chem.* **2001**, *276*, 31602-31612.

[138] P. Mehta, R.D. Cummings, R.P. McEver, *J. Biol. Chem.* **1998**, *273*, 32506-32513.

[139] M.W. Nicholson, A.N. Barclay, M.S. Singer, S.D. Rosen, P.A. van der Merwe, *J. Biol. Chem.* **1998**, *273*, 763-770.

[140] A.F. Williams, *Nature* **1991**, *352*, 473-474.

[141] J.Y. Ramphal, Z.L. Zheng, C. Perez, L.E. Walker, S.A. DeFrees, F.A. Gaeta, *J. Med. Chem.* **1994**, *37*, 3459-3463.

[142] B.K. Brandley, M. Kiso, S. Abbas, P. Nikrad, O. Srivasatava, C. Foxall, Y. Oda, A. Hasegawa, *Glycobiology* **1993**, *3*, 633-639.

[143] W. Stahl, U. Sprengard, G. Kretzschmar, H. Kunz, *Angew. Chem. Int. Ed. Engl.* **1994**, *22*, 2096-2098.

[144] H. Ohmoto, K. Nakamura, T. Inoue, N. Kondo, K. Yoshino, H. Kondo, H. Ishida, M. Kiso, A. Hasegawa, *J. Med. Chem.* **1996**, *39*, 1339-1343.

[145] D. Tyrell, P. James, N. Rao, C. Foxall, S. Abbas, F. Dasgupta, M. Nashed, A. Hasegawa, M. Kiso, D. Asa, J. Kidd, B.K. Brandley, *Proc. Natl. Acad. Sci. USA* **1991**, *88*, 10372-10376.

[146] S.A. DeFrees, F.A. Gaeta, Y.C. Lin, Y. Ichikawa, C.H. Wong, *J. Am. Chem. Soc.* **1993**, *115*, 7549-7550.

[147] Y. Hiramatsu, H. Tsujishita, H. Kondo, *J. Med. Chem.* **1996**, *39*, 4547-4553.

[148] Y. Wada, T. Saito, N. Matsuda, H. Ihmoto, K. Yoshino, M. Ohashi, H. Kondo, H. Ishida, M. Kiso, A. Hasegawa, *J. Med. Chem.* **1996**, *39*, 2055-2059.

[149] N. Kaila, B. E. I. V. Thomas, *Med. Res. Rev.* **2002**, *22*, 566-601.

[150] T.J. Rutherford, D.G. Spackman, P.J. Simpson, S.W. Homans, *Glycobiology* **1994**, *4*, 59-68.

[151] R. Harris, G. R. Kiddle, R. A. Field, M. J. Milton, B. Ernst, J. L. Magnani, S. W. Homans, *J. Am. Chem. Soc.* **1999**, *121*, 2546-2551.

[152] Y. Ichikawa, Y.C. Lin, D.P. Dumas, G.J. Shen, E. Garcia-Junceda, M.A. Williams, R. Bayer, C. Ketcham, L.E. Walker, J. Paulson, C.H. Wong, *J. Am. Chem. Soc.* **1992**, *114*, 9283-9298.

[153] L. Poppe, G.S. Brown, J.S. Philo, P. Nikrad, B.H. Shah, *J. Am. Chem. Soc.* **1997**, *119*, 1727-1736.

[154] K. Veluraja, C.J. Margulis, *J. of Biomol. Struct. & Dynamics* **2005**, *23*, 101-111.

[155] S. Perez, N. Mouhous-Riou, N.E Nifant'ev, Y.E. Tsvetkov, B. Bachet, A. Imberty, *Glycobiology* **1996**, *6*, 537-542.

[156] K.E. Miller, C. Mukhopadhyay, P. Cagas, C.A. Bush, Biochemistry **1992**, *31*, 6703-6709.

[157] R.M. Cooke, *Biochemistry* **1994**, *33*, 10591-10596.

[158] I. Tvaroska, M. Hricovini, E. Petrakova, *Carbohydr. Res* **1989**, *189*, 359-362.

[159] B. Bose, S. Zhao, R. Stenutz, F. Cloran, P. B. Bondo, G. Bondo, B. Hertz, I. Carmichael, A. S. Serianni, *J. Am. Chem. Soc.* **1998**, *120*, 11158-11173.

[160] M.J. Milton, R. Harris, M.A. Probert, R.A. Field, S.W. Homans, *Glycobiology* **1998**, *8*, 147-153.

[161] H.C. Kolb, B. Ernst, *Chem. Eur. J.* **1997**, *3*, 1571-1578.

[162] R.M. Cooke, *Biochemistry* **1994**, *33*, 10591-10596.

[163] P. Hensley, P. J. McDevitt, I. Brooks, J. J. Trill, J. A. Feild, D. E. McNulty, J. R. Connor, D. E. Griswold, N. V. Kumar, K. D. Kopple, *J. Biol. Chem.* **1994**, *269*, 23949-23958.

[164] K. Scheffler, J. R. Brisson, R. Weisemann, J. L. Magnani, W. T. Wong, B. Ernst, T. Peters, *J. Biomolecul. NMR* **1997**, *9*, 423-436.

[165] K. Scheffler, B. Ernst, A. Katopodis, J. L. Magnani, W. T. Wang, R. Weisemann, T. Peters, *Ang. Chem. Int. Ed.* **1995**, *34*, 1841-1844.

[166] W.S. Somers, J. Tang, G.D. Shaw, R.T. Camphausen, *Cell* **2000**, *103*, 467-479.

[167] R. Harris, G. R. Kiddle, R. A. Field, M. J. Milton, B. Ernst, J. L. Magnani, S. W. Homans, *J. Am. Chem. Soc.* **1999**, *121*, 2546-2551.

[168] H. C. Kolb, B. Ernst, *Pur. Appl. Chem.* **1997**, *69*, 1879-1884.

[169] F. Mohamadi, N.G.J Richards, W.C. Guida, R. Liskamp, M. Lipton, C. Caufield, G. Chang, T. Hendrickson, W.C. Still, *Journal of Comput. Chem.* **1990**, *11*, 440-467.

[170] H. Senderowitz, C. Parish, W. C. Still, *J. Am. Chem. Soc.* **1996**, *118*, 2078-2086.

[171] W. C. Still, A. Tempczyk, R. C. Hawley, T. Hendrickson, *J. Am. Chem. Soc.* **1990**, *112*, 6127-6129.

[172] K.K.-S. Ng, W.I. Weis, *Biochemistry* **1997**, *36*, 979-988.

[173] B. J. Graves, R. L. Crowther, C. Chandran, J. M. Rumberger, S. Li, K-S. Huang, D. H. Presky, P. C. Familletti, B. A. Wolitzky, B.K Burns, *Nature* **1994**, *367*, 532-538.

[174] T.P. Kogan, B.M. Revelle, S. Tapp, D. Scott, P.J. Beck, *J. Biol. Chem.* **1995**, *270*, 14047-14055.

[175] B. Ernst, Z. Dragic, S. Marti, C. Müller, B. Wagner, W. Jahnke, J.L. Magnani, K.E. Norman, R. Oehrlein, T. Peters, H.C. Kolb, *Chimia* **2001**, *55*, 268-274.

[176] J.Y. Ramphal, M. Hiroshige, B. Lou, J.J. Gaudino, M. Hayashi, S.M. Chen, L.C. Chiang, F.A. Gaeta, S.A. DeFrees, *J. Med. Chem.* **1996**, *39*, 1357.

[177] A. Seppo, J.P. Turunen, L. Penttila, A. Keane, O. Renkonen, R. Renkonen, *Glycobiology* **1996**, *6*, 182.

[178] O. Renkonen, R. Renkonen, S. Topipila, L. Penttila, H. Salminen, J.P. Turunen, *Glycobiology* **1997**, *7*, 453.

[179] C.-H. Lin, M. Shimazaki, C.H. Wong, M. Koketsu, L.R. Juneja, J.M. Kim, *Bioorg. Med. Chem. Lett.* **1995**, *3*, 1625.

[180] T. Ikeda, T. Kajimoto, H. Kondo, C.H. Wong, *Bioorg. Med. Chem. Lett.* **1997**, *4*, 22485.

[181] L.L. Kiessling, N. Pohl, *Chem. Biol.* **1996**, *3*, 71.

[182] E.E. Simanek, G.J. McGarvey, J.A. Jablonowski, C.H. Wong, *Chem. Rev.* **1998**, *98*, 833-862.

[183] N. Sakagami, K. Horie, K. Nakamoto, T. Kawaguchi, H. Hamana, *Bioorg. Med. Chem. Lett.* **1998**, *8*, 2783-2786.

[184] P. Sears, C.H. Wong, *Angew. Chem. Int. Ed.* **1999**, *38*, 2300-2324.

[185] G. Thoma, J.T. Patton, J. Magnani, B. Ernst, R. Oehrlein, R.O. Duthaler, *J. Am. Chem. Soc.* **1999**, *12*, 5919-5929.

[186] R. Stahn, H. Schaeffer, F. Kernchen, J. Schreiber, *Glycobiology* **1998**, *8*, 311-319.

[187] R. Roy, *Curr. Opin. Struct. Biol.* **1996**, *6*, 692-702.

[188] M. Mammen, S.K. Choi, G.M. Whitesides, *Angew. Chem. Int. Ed.* **1998**, *37*, 2754-2794.

[189] J.M. Gardiner, *Exp. Opin. Invest. Drugs* **1998**, *7*, 405-410.

[190] D.D. Manning, C.R. Bertozzi, S.D. Rosen, L.L. Kiessling, *Tetrahedron Lett.* **1996**, *37*, 1953-1956.

[191] N. Imazaki, H. Koike, H. Miyauchi, M. Hayashi, *Bioorg. Med. Chem. Lett.* **1996**, *6*, 2043-2048.

[192] G. Thoma, F. Schwarzenbach, R.O. Duthaler, *J. Org. Chem.* **1996**, *61*, 514-524.

[193] S. Hanessian, H.K. Huynh, G.V. Reddy, G. McNaughton-Smith, B. Ernst, H.C. Kolb, J. Magnani, C. Sweeley, *Bioorg. Med. Chem. Lett.* **1998**, *8*, 2803-2808.

[194] A. Töpfer, G. Kretzschmar, E. Bartnik, *Tetrahedron Lett.* **1995**, *36*, 9161-9164.

[195] M.J. Bamford, M. Bird, P.M. Gore, D.S. Holmes, R. Priest, J.C. Prodger, V. Saez, *Bioorg. Med. Chem. Lett.* **1996**, *6*, 239-244.

[196] G. Thoma, J.L. Magnani, J.T. Patton, B. Ernst, W. Jahnke, *Angew. Chem. Int. Ed.* **2001**, *40*, 1941-1945.

[197] G. Thoma, R. Bänteli, W. Jahnke, J. Magnani, J.T. Patton, *Angew. Chem.* **2001**, *113*, 3756-3759.

[198] S. Hanessian, H. Prabhanjan, *Synlett* **1994**, 868-870.

[199] A.A. Birkbeck, S.V. Ley, J.C. Prodger, *Bioorg. Med. Chem. Lett.* **1995**, *5*, 2637-2642.

[200] N. Kaila, H.-A. Yu, Y. Xiang, *Tetrahedron Lett.* **1995**, *36*, 5503-5506.

[201] A. Töpfer, G. Kretzschmar, S. Schuth, M. Sonnentag, *Bioorg. Med. Chem. Lett.* **1997**, *7*, 1317-1322.

[202] R. Bänteli, B. Ernst, *Tetrahedron Lett.* **1997**, *38*, 4059-4062.

[203] A. Liu, K. Dillon, R.M. Campbell, D.C. Cox, D.M. Huryn, *Tetrahedron Lett.* **1996**, *37*, 3785-3788.

[204] B. Dupre, H. Bui, I.L. Scott, R.V. Market, K.M. Keller, P.J. Beck, T.P. Kogan, *Bioorg. Med. Chem. Lett.* **1996**, *6*, 569-572.

[205] T.P. Kogan, B. Dupre, K.M. Keller, I.L. Scott, H. Bui, R.V. Market, P.J. Beck, J.A. Voytus, B.M. Revelle, D. Scott, *J. Med. Chem.* **1995**, *38*, 4976-4984.

[206] L. Pardella, *Anti-inflammatory immunomodulatory Invest. Drugs* **1999**, *1*, 56-60.

[207] C.-H. Wong, F. Moris-Varas, S.-C. Hung, T.G. Marron, C.-C. Lin, K.W. Gong, G. Weitz-Schmidt, *J. Am. Chem. Soc.* **1997**, *119*, 8152-8158.

[208] T.G. Marron, T.J. Woltering, G. Weitz-Schmidt, C.-H. Wong, *Tetrahedron Lett.* **1996**, *37*, 9037-9040.

[209] R. Wang, C.-H. Wong, *Tetrahedron Lett.* **1996**, *37*, 5427-5430.

[210] S.H. Wu, M. Shimazaki, C.C. Lin, L. Quiao, W.J. Moree, G. Weitz-Schmidt, C.H. Wong, *Angew. Chem. Int. Ed.* **1996**, *35*, 88-90.

[211] T.J. Woltering, G. Weitz-Schmidt, C.H. Wong, *Tetrahedron Lett.* **1996**, *37*, 9033.

[212] M.W. Cappi, W.J. Moree, L. Quiao, T.G. Marron, G. Weitz-Schmidt, C.H. Wong, *Bioorg. Med. Chem. Lett.* **1997**, *5*, 283-296.

[213] Y. Hiramatsu, T. Tsukida, Y. Nakai, Y. Inoue, H. Kondo, *J. Med. Chem.* **2000**, *43*, 1476-1483.

[214] B.N.N. Rao, M.B. Anderson, J.H. Musser, J.H. Gilbert, M.E. Schaefer, C. Foxall, B.K. Brandle, *J. Biol. Chem.* **1994**, *269*, 19663-19666.

[215] B. Ernst, H. Kolb, O. Schwardt, *Carbohydrate Mimetics in Drug Discovery* in *The Organic Chemistry of Sugars* **2006**, CRC Press, 803-861.

[216] H.A. Scheraga, *Chem. Rev.* **1971**, *71*, 195-217.

[217] A.K. Ghose, G.M. Crippen, G.R. Revankar, D.F. Smee, P.A. McKernan, R.K. Robin, *J. Med. Chem.* **1989**, *32*, 746-756.

[218] A.E. Howald, P.A. Kollman, *J. Med. Chem.* **1988**, *31*, 1669-1675.

[219] M. Lipton, W.C. Still, *J. Comput. Chem.* **1988**, *9*, 343-355.

[220] R.A. Dammkoehler, S.F. Karasek, E.F.B. Shands, G.R. Marshall, *J. Comput.-Aided Mol. Design* **1989**, *3*, 3-21.

[221] M. Saunders, *J. Am. Chem. Soc.* **1987**, *109*, 3150-3152.

[222] M. Saunders, *J. Comput. Chem.* **1989**, *10*, 203-208.

[223] D.M. Ferguson, D.J. Raber, *J. Am. Chem. Soc.* **1989**, *111*, 4371-4378.

[224] G. Chang, W.C. Guida, W.C. Still, *J. Am. Chem. Soc.* **1989**, *111*, 4379-4386.

[225] M. Saunders, K.N. Houk, Y.-D. Wu, W.C. Still, M. Lipton, G. Chang, W.C. Guida *J. Am. Chem. Soc.* **1990**, *112*, 1419-1427.

[226] A.K. Ghose, E.P. Jaeger, P.J. Kowalczyk, M.L. Peterson, A.M. Treasurywala, *J. Comput. Chem.* **1993**, *14*, 1050-1065.

[227] H.-J. Böhm, G. Klebe, T. Lorenz, T. Mietzner, L. Siggel, *J. Comput. Chem.* **1990**, *11*, 1021-1028.

[228] R. Taylor, G.W. Mullier, G.J. Sexton, *J. Mol. Graphics.* **1992**, *10*, 152-160.

[229] P.S. Shenkin, D.Q. McDonald, *J. Comput. Chem.* **1994**, *15*, 899-916.

[230] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, E. Teller, *J. Chem. Phys.* **1953**, *32*, 1087-1092.

[231] J.M. Goodman, W.C. Still, *J. Comput. Chem.* **1991**, *12*, 1110-1117.

[232] S. Kirkpatrick, C.D. Gelatt, M.P. Vecchi, *Science* **1983**, *220*, 671-680.

[233] Laine, R. A. *Glycobiology* **1994**, *4*, 759.

[234] E.W. Wooten, C.J. Edge, R. Bazzo, R.A. Dwek, T.W. Rademacher, *Carbohydr. Res.* **1990**, *203*, 13-17.

[235] A. Imberty, S. Perez, *Chem. Rev.* **2000**, *100*, 4567-4588.

[236] R.J. Woods, *Glycoconjugate J.* **1998**, *15*, 209-216.

[237] J.L. Jenkins, R.Y.T. Kao, R. Shapiro, *Proteins* **2003**, *51*, 81-93.

[238] C.M. Venkatachalam, X. Jiang, T. Oldfield, M. Waldman, *J. Mol. Graph. Model.* **2003**, *21*, 289-307.

[239] H.J. Park, J. Yoo, Y.K. Kim, J.W. Han, H.W. Lee, In 224th ACS National Meeting; American Chemical Society, Washington, D. C: Boston, MA, United States, 2002.

[240] S. Grueneberg, M.T. Stubbs, G. Klebe, *J. Med. Chem.* **2002**, *45*, 3588-3602.

[241] T.N. Doman, S.L. McGovern, B.J. Witherbee, T.P. Kasten, R. Kurumbail, W.C. Stallings, D.T. Connolly, B.K. Shoichet, *J. Med. Chem.* **2002**, *45*, 2213-2221.

[242] G. Rastelli, A.M. Ferrari, L. Costanitino, M.C. Gamberini, *Bioorg. Med. Chem.* **2002**, *10*, 1437-1450.

[243] P. Cozzini, T. Dottorini, *Eur. J. Med. Chem.* **2004**, *39*, 601-609.

[244] T. Hou, W. Zhang, X. Xu, *J. Comput. Aided Mol. Des.* **2002**, *16*, 27-41.

[245] Y.S. Shiau, P.T. Huang, H.H. Liou, Y.C. Liaw, Y.Y Shiau, K.L Lou, *Chem. Res. Toxicol.* **2003**, *16*, 1217-1225.

[246] H.F. Chen, Q. Li, X. Yao, B.T. Fan, S.G. Yuan, A. Panaye, J.P. Doucet, J. P. *QSAR Comb. Sci.* **2003**, *22*, 604-613.

[247] A.C. Illapakurthy, C. Ashok, Y.A. Yogesh, B.A. Avery, M.A. Avery, C.M. Wyandt, *J. Pharm. Sci.* **2003**, *92*, 649-655.

[248] A. Vedani, D.W. Huhta, *J. Am. Chem. Soc.* **1990**, 112, 4759-4767.

[249] I.D. Kuntz, J.M. Blaney, S.J. Oatley, R. Langringe, T.E. Ferrin, *J. Mol. Biol*. **1982**, *161*, 269-288.

[250] D. Fischer, R. Norel, H. Wolfson, R. Nussinov, *Proteins* **1993**, *16*, 278-292.

[251] D. Fischer, S.L. Lin, H.L. Wolfson, R. Nussinov, *J. Mol. Biol.* **1995**, *248*, 459-477.

[252] H.-J. Boehm, *J. Comput. Aided Mol. Des.* **1992**, *6*, 593-606.

[253] H.-J. Boehm, *J. Comput. Aided Mol. Des.* **1992**, *6,* 61-78.

[254] S.K. Kearsley, D.J. Underwood, R.P. Sheridan, M.D. Miller, J*. Comput. Aided Mol. Des.* **1994**, *8*, 565-482.

[255] D.M. Lorber, B.K. Shoichet, *Protein Science* **1998**, *7*, 938-950.

[256] M. Rarey, B. Kramer, T, Lengauer, G. Klebe, *J. Mol. Biol.* **1996**, *261*, 470-489.

[257] M. Rarey, B. Kramer, T. Lengauer, *J. Comput. Aided Mol. Des.* **1997**, *11*, 369-384.

[258] D.S. Goodsell, A.J. Olson, *Proteins* **1990**, *8*, 195-202.

[259] J.A. Given, M.K. Gilson, *Proteins* **1998**, *33*, 475-495.

[260] J.Y. Trosset, H.A. Scheraga, *J. Comput. Chem.* **1999**, *20*, 412-427.

[261] J.Y. Trosset, H.A. Scheraga, *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 8011-8015.

[262] J.Y. Trosset, H.A. Scheraga, J*. Comput. Chem.* **1999**, *20*, 244-252.

[263] G. Jones, P. Willett, R.C. Glen, A.R. Leach, R. Taylor, *J. Mol. Biol.* **1997**, *267*, 727-748.

[264] G. Jones, P. Willett, R.C. Glen, *J. Comput. Aided Mol. Des.* **1995**, *9*, 532-549.

[265] T. Schulz-Gasch, M. Stahl, *J. Mol. Model.* **2003**, *9*, 47-57.

[266] M. Kontoyianni, L.M. McClellan, G.S. Sokol, *J. Med. Chem.* **2004**, *47*, 558-565.

[267] G. Wu, D.H. Robertson, C.L. 3rd Brooks, M. Vieth, *J. Comput. Chem.* **2003**, *24*, 1549-1562.

[268] E. Perola, W.P. Walters, P.S. Charifson, *Proteins* **2004**, *56*, 235-249.

[269] I. Muegge, M. Rarey, *Rev. Comput. Chem.* **2001**, *17*, 1-60.

[270] A. Vedani, H. Brien, M. Dobler, H. Dollinger, D.R. McMaster, *J. Med. Chem.* **2000**, *43*, 4416-4427.

[271] A. Vedani, M. Dobler, *J. Med. Chem.* **2002**, *45*, 2139-2149.

[272] A. Vedani, M. Dobler, M.A. Lill, *J. Med. Chem.* **2005**, *48*, 3700-3703.

[273] R.D. Taylor, P.J. Jewsbury, J.W. Essex, *J. Comput. Aided Mol. Des.* **2002**, *16*, 151-166.

[274] I. Halperin, B. Ma, H. Wolfson, R. Nussinov, *Proteins* **2002**, *47*, 409-443.

[275] J.R.H. Tame, *J. Comput. Aided Mol. Des.* **1999**, *13*, 99-108.

[276] H.-J Boehm, *J. Comput. Aided Mol. Des.* **1994**, *8*, 243-256.

[277] M.D. Eldridge, C.W. Murray, T.R. Auton, G.V. Paolini, R.P. Mee, *J. Comput. Aided Mol. Des.* **1997**, *11*, 425-445.

[278] I. Muegge, Y.C. Martin, *J. Med. Chem.* **1999**, *42*, 791-804.

[279] H. Gohlke, M. Hendlich, G. Klebe, *J. Mol. Biol.* **2000**, *295*, 337-356.

[280] C. Bissantz, G. Folkers, D. Rognan, *J. Med. Chem.* **2000**, *43*, 4759-4767.

[281] M. Stahl, M. Rarey, *J. Med. Chem.* **2001**, *44*, 1035-1042.

[282] P.S. Charifson, J.J. Corkery M.A. Murcko, W.P. Walters, *J. Med. Chem.* **1999**, *42*, 5100-5109.

[283] M.L. Verdonk, V. Berdini, M.J. Hartshorn, W.T.M. Mooij, C.W. Murray, R.D. Taylor, P. Watson, *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 793-806.

[284] Y. Pan, N. Huang, S. Cho, A.D.J. MacKerell, *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 267-272.

[285] P. Pospisil, P. Ballmer, L. Scapozza, G. Folkers, *J. Recept. Signal Transduct. Res.* **2003**, *23*, 361-371.

[286] Ramsden, C.A., *Quantitative Drug Design* (Comprehensive Medicinal Chemistry) (Vol. 4) **1990**, Fergamon Press.

[287] H. Kubinyi, *QSAR: Hansch Analysis and Related Approacbes* **1993**, VCH.

[288] C. Hansch, A. Leo, *Exploring QSAR. Fundamentals and Applications in Chemistry and Biology* **1995**, American Chemical Society.

[289] C. Hansch, A. Leo, D. Hoekman, *Exploring QSAR: Hydrophobic, Electronic, and Steric Constants* **1995**, American Chemical Society.

[290] H. van de Waterbeemd, *Chemometric Methods in Molecular Design* **1995**, VCH.

[291] H. van de Waterbeemd, *Advanced Computer-Assisted Techniques in Drug Discovery* **1995**, VCH.

[292] H. Kubinyi, in Burger's Medicinal Chemistry (Vol. 1, 5th Ed.) **1995**, Ed. Wiley, 497-571.

[293] F. Sanz, J. Giraldo, F. Manut, *QSAR and Molecular Modeling: Concepts, Computational Tools and Biological Applications* **1994**, Prous Science.

[294] V. Pliska, B. Testa, H. van de Waterbeemd, *Lipophilicity in Drug Action and Toxicology* **1996**, VCH.

[295] H. van de Waterbeemd, *Structure-Property Correlations in Drug Research* **1996**, Academic Press.

[296] H. van de Waterbeemd, *The Practice of Medicinal Chemistry* **1996**, Academic Press, 367-389.

[297] H.-J. Boehm, G. Klebe, H. Kubinyi, *Wirkstoffdesign* **1996**, Spektrum Akademischer Verlag, 363-436.

[298] H. van de Waterbeemd, B. Testa, G. Folkers, *Computer-Assisted Lead Finding and Optimization* **1996**, VCH.

[299] S.M. Jr Free, J.W. Wilson, *J. Med. Chem.* **1964**, *7*, 395-399.

[300] C. Hansch, T. Fujita, *J. Am. Chem. Soc.* **1964**, *86*, 1616-1626.

[301] R.D. Cramer, D.E. Patterson III, J.D. Bunce, *J. Am. Chem. Soc.* **1988**, *110*, 5959-5967.

[302] H. Kubinyi, *3D QSAR in Drug Design. Theory, Methods and Applications* **1993**, ESCOM Science Publishers.

[303] H. Kubinyi, G. Folkers, Y.C. Martin, *3D QSAR in Drug Design: Ligand-Protein Interactiond and Molecular Similarity (Vol. 2) and Recent Advances (Vol. 3)* **1997**, Kluwer Academic Publishers.

[304] U. Norinder, *Perspectives in Drug Discovery and Design* **1998**, *12*, 25-39.

[305] K.H. Kim, G. Greco, E. Novellino, *Perspectives in Drug Discovery and Design* **1998**, *12*, 257-315.

[306] G. Folkers, A. Merz, D. Rognan, *CoMFA: Scope and Limitations* in H. Kubinyi, *3D QSAR in Drug Design: Theory, Methods and Applications* **1993**, ESCOM Science Publishers, 583-618.

[307] G. Klebe, U. Abraham, T. Mietzner, *J. Med. Chem.* **1994**, *37*, 4130-4146.

[308] G. Klebe, U. Abraham, *J. Comput. Aided. Mol. Design* **1999**, *13*, 1-10.

[309] Boehm M., J. Stuerzenbecher, G. Klebe, *J. Med. Chem.* **1999**, *42*, 458-477.

[310] Boehm M., G. Klebe, *J. Med. Chem.* **2002**, *45*, 1585-1597.

[311] M. Baroni, G. Costantino, G. Cruciani, D. Riganelli, R. Valigi, S. Clementi, *Quant. Struct.-Act. Relat.* **1993**, *12*, 9-20.

[312] G. Cruciani, K.A. Watson, *J. Med. Chem.* **1994**, *37*, 2589-2601.

[313] E. Walters, *Perspectives in Drug Discovery and Design* **1998**, *12*, 159-166.

[314] A.N. Jain, K. Kolie, D.J. Chapman, *J. Med. Chem.* **1994**, *37*, 2315-2327.

[315] M.A. Lill, A. Vedani, M. Dobler, *J. Med. Chem.* **2004**, *47*, 6174-6186.

[316] M.A. Lill, F. Winiger, A. Vedani, B. Ernst, *J. Med. Chem.* **2005**, *48*, 5666-5674.

[317] G. Schneider, U. Fechner, *Nature Review Drug Design* **2005**, *4*, 649-663.

[318] R.S. Bohacek, C. McMartin, *J. Am. Chem. Soc.* **1994**, *116*, 5560-5571.

[319] C. McMartin, R.S. Bohacek, *J. Comput. Aided. Mol. Design* **1997**, *11*, 333-344.

[320] N.P Todorov, P.M Dean, *J. Comput. Aided. Mol. Design* **1997**, *11*, 175-192.

[321] N.P Todorov, P.M Dean, *J. Comput. Aided. Mol. Design* **1998**, *12*, 335-350.

[322] D.A. Pearlman, D.A. Case, J.W. Caldwell, W.R. Ross, T.E. Cheatham, III, S. DeBolt, D. Ferguson, G. Seibel, P. Kollman, *Comp. Phys. Commun.* **1995,** *91*, 1-41.

[323] W.F. van Gunsteren, H.J.C. Berendsen, *Angew. Chemie* **1990**, *102*, 1020-1055.

[324] J. Wang, W. Wang, P.A. Kollman, D.A. Case, *J. Chem. Inf. Comput. Sci.,* (submitted).

[325] D.A. Case, T.E. Cheatham, III, T. Darden, H. Gohlke, R. Luo, K.M. Merz, Jr., A. Onufriev, C. Simmerling, B. Wang, R. Woods, *J. Computat. Chem.* **2005** , *26***,** 1668-1688.

[326] C.J. Cramer, D.G. Truhlar, *J. Comp.-Aided Mol. Design* **1992**, *6*, 629-666.

[327] J.W. Storer, D.J. Giesen, C.J. Cramer, D.G. Truhlar, *J. Comput.-Aid. Mol. Des.* **1995**, *9*, 87.

[328] Biographics Laboratory 3R, Friedensgasse 35, CH-4056 Basel, Switzerland.

[329] J. J. P. Stewart, *J. Comp.-Aided Mol. Design* **1990**, *4*, 1-105.

[330] A. Jakalian, D.B. Jack, C.I. Bayly, *J. Comput. Chem.* **2002**, *23*, 1623-1641.

[331] M.J.S. Dewar, E.G. Zoebisch, E.F. Healy, J.J.P. Stewart, *J. Am. Chem. Soc.* **1985**, *107*, 3902-3909.

[332] http://plotamber.sourceforge.net/

[333] P. Zbinden, M. Dobler, G. Folkers, A. Vedani, *Quant. Struct.-Act. Relat.* **1998**, *17*, 122-130.

[334] W.L. DeLano, *The PyMOL Molecular Graphics System* **2002**, DeLano Scientific, San Carlos, CA, USA. http://www.pymol.org

[335] http://www.biograf.ch/PDFS/Raptor.pdf

[336] S. Schmid, Institute of Molecular Pharmacy, University of Basel (unpublished).

[337] S. Schmid, Institute of Molecular Pharmacy, University of Basel (unpublished).

[338] D.E. Walters, R.M. Hinds, *J. Med. Chem.* **1994**, *37*, 2527-2536.

[339] D. Rogers, A.J. Hopfinger, *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 854-866.

[340] A. Vedani, J.D. Dunitz, *J. Am. Chem. Soc.* **1985**, *107*, 7653-7658.

[341] http://www.ks.uiuc.edu/Research/vmd

[342] W. Humphrey, A. Dalke, K. Schulten, *J. Mol. Graph.* **1996**, *14*, 33-38.

[343] R. Weimar, R.J. Woods, *Combining NMR and Simulation Methods in Oligosaccharide Conformational Analysis*, in *NMR Spectroscopy of Glycoconjugates* **2003**, VCH, 111-142.

[344] R.J. Woods, R.A. Dwek, C.J. Edge, B. Fraser-Reid, *J. Phys. Chem.* **1995**, *99*, 3832-3846.

[345] D.A. Case, D.A. Pearlman, J.W. Caldwell, T.E. Cheatham III, W.S. Ross, C.L. Simmerling, T.A. Darden, K.M. Merz, R.V. Stanton, A.L. Cheng, J.J. Vincent, M. Crowley, V. Tsui, R.J. Radmer, Y. Duan, J. Pitera, I. Massova, G.L. Seibel, U.C. Singh, P.K. Weiner, P.A. Kollman, *Solvation algorithm of Leap* in the *AMBER 6 Manual* 1996, University of California at San Francisco.

[346] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne, *Nucleic Acids Research* **2000**, *28,* 235-242.

[347] A. Vedani, personal communication.

[348] C. Bissantz, G. Folkers, D. Rognan, *J. Med. Chem.* **2000**, *43*, 4759-4767.

[349] B. Ernst, personal communication

[350] X. Zhang, D.F. Bogorin, V.T. Moy, *ChemPhysChem* **2004**, *5*, 175-182.

[351] http://www.biograf.ch/PDFS/Quasar.pdf

[352] http://www.biograf.ch/PDFS/Yeti.pdf

[353] H.J. Gambius, *Naturwissenschafter* **2000**, *87*, 108-121.

[354] C. Mueller, Dissertation: Selectin Antagonists: Synthesis and Conformational Comparison of C- and O-glycosidic Tetrasaccharide Mimetics related to Sialyl Lewis[x], **2005**, Institute of Molecular Pharmacy, University of Basel

[355] A. Titz, ongoing dissertation, Institute of Molecular Pharmacy, University of Basel.

[356] H. Gohlke, G. Klebe, *Angew. Chem.* **2002**, *114*, 2764-2798.

[357] E. Freire, *Thermodynamics in drug design. High affinity and selectivity* in *Proceeding of The Chemical Theatre of Biological Systems* **2004**.

[358] L. Tschopp, ongoing dissertation, Institute of Molecular Pharmacy, University of Basel.

[359] R. Bänteli, B. Ernst, *Bio. Med. Chem. Lett*. **2001**, *11*, 459-462.

[360] D. Schwizer, ongoing dissertation, Institute of Molecular Pharmacy, University of Basel.

[361] A.M. Davis, S.J. Teague, *Angew. Chem. Int. Ed. Engl.* **1999**, *38*, 736-749.

[362] C.A. Lipinski*,* F. Lombardo, B.W. Dominy, P.J. Feeney, *Advanced Drug Delivery Reviews* **1997**, *23*, 3-25.

[363] T.I. Oprea, *Journal of Computer-Aided Molecular Design* **2002**, *16*, 325-334.

8 Curriculum Vitae

Michele Porro

Birmannsgasse 46	Via R. Simen 9	☎ (Private): +41 (61) 272.42.58
4055 Basel	6830 Chiasso	☎ (Office): +41 (61) 267.15.63
Switzerland	Switzerland	☎ (Mobile): +41 (78) 825.46.47
		E-mail: michele.porro@unibas.ch

PERSONAL STATUS

Date of Birth: October 05, 1977
Nationality: Swiss
Marital Status: Single
Sex: Male

EDUCATION

July 2002 - present	**PhD student in Molecular Modeling, Institute of Molecular Pharmacy, University of Basel, Basel, Switzerland** Thesis Title: Design, Analysis and *in silico* Evaluation of E-Selectin Antagonists Supervisors: Prof. Dr. B. Ernst and Prof. Dr. A. Vedani
July 2002 - present	**Graduate student program of the Pharmacenter, Basel-Zurich, Switzerland**
September 2005- October 2005	**Course in "Key issues in Drug Discovery & Development"** An 8 days course that gave an overview of today's drug development strategies, Pharmacenter, Basel-Zurich, Switzerland
April 2005	**Computer Aided Molecular Design Group, Institute of Pharmacy, University of Innsbruck, Innsbruck, Austria** Two-week internship with Prof. Dr. T. Langer to learn more about pharmacophore modeling and about Catalyst software application
September 2004	**EMBO-Course, Shanghai, China** Title: Methods for protein simulation and drug design Financial support from EMBO (1 of the 6 foreign students invited after selection)
May 2003	**4th European Workshop on Drug Design, University of Siena, Certosa di Pontignano, Italy,** The participation was financed by the University of Basel (Resort Nachwuchsförderung).
1996 - 2001	**Master degree in Pharmacy,** (Eidg. Dipl. Apotheker)**, University of Basel, Basel, Switzerland,** Received 5.5/6.0 theoretical, 6.0/6.0 practical for diploma

February 2001 - July 2001	**Master Thesis, Institute of Molecular Pharmacy, University of Basel, Basel, Switzerland** Title: Ein QSAR-Modell zur Entwicklung von hochaffinen Liganden für den Asialoglykoproteinrezeptor Supervisor: Prof Dr. A. Vedani

PROFESSIONAL EXPERIENCE

May 2005 – September 2005	**Diploma Work Supervisor, Institute of Molecular Pharmacy, University of Basel, Basel, Switzerland** Leading, coordinating and supporting students in their diploma work Project: *In Silico* Charakterisierung von E-Selektin Antagonisten
May 2004 - September 2004	**Diploma Work Supervisor, Institute of Molecular Pharmacy, University of Basel, Basel, Switzerland** Leading, coordinating and supporting students in their diploma work Project: A new QSAR Model for E-Selectin.
July 2002 - present	**Assistant, Institute of Molecular Pharmacy, University of Basel, Basel, Switzerland** Coordinate monthly project meetings attended by chemists, biologists and modelers Lecture graduate students on several topics related to drug design and management of laboratory sessions in chemical molecular modeling
August 1999 - present	**Pharmacist, Local Pharmacies, Novazzano and Basel, Switzerland** Performed one-year internship to fulfill diploma requirements, and worked in pharmacies during the studies
August 1995	**Trainee, PAMP S.A, Castel S. Pietro, Switzerland** Worked as an assistant in the analytic department (gold quality analysis)

PUBLICATIONS, POSTERS AND ORAL PRESENTATIONS

D. Schwizer, H. Gathje, S. Kelm, M.Porro, O. Schwardt, and B. Ernst, Antagonists of the Myelin-associated Glycoprotein: a new Class of Tetrasaccharide Mimics, **Biorg. & Med. Chem** (*in press*)
S. Shelke, D. S. Strasser, B. Cutting, M.Porro, G. Gao, H. Gathije, S. Kelm, O. Schwardt and B. Ernst, Exploring Sialic Acid Derivatives as Myelin-associated Glycoprotein Ligands (*in preparation*)
M.Porro, M. Schmid and B. Ernst, A new Structure-Activity Model for E-Selectin (*in preparation)*
M.Porro, M. Schmid and B. Ernst, A new Structure-Activity Model for E-Selectin, Poster and oral presentation at the EMBO-Course, Shanghai, September 2004
M.Porro and B. Ernst, Molecular Dynamics of Selectin Ligand Interaction, Poster at the Pharmaday ETH-UniBS, Zurich, February 2004
M.Porro and B. Ernst, Carbohydrates-Lectin Interactions, Poster at the 4th European Workshop on Drug Design, Siena, May 2003

MOLECULAR MODELING AND OTHER SOFTWARE KNOWLEDGE

MacroModel 5.0-8.0: primarily used to perform minimizations and conformational searches
Yeti 6.6: Docking software, particularly suited to work with metallo-proteins
Quasar 5.0 and Raptor 2.0: to develop QSAR-Models
PrGen 2.1: to perform receptor mediated alignments
Amber 7-8: particularly suited for running Minimizations and MD with carbohydrates
        (GLYCAM-force field)
Mopac 7.0, Autodock 3.0, Sybyl 6.9, QXP, Flo+, Growmol: routine use
Unix, OSX, Windows XP, Redhat 8.0, Catalyst 4.9, Grid 22a: basic knowledge
Powerpoint, Excel, Word: good knowledge

EXTRACURRICULAR ACTIVITIES

July 2002 - present	Representative of the PhD-students in the University Commission devoted to the reform of the Pharmacy studies
May 1999 - October 2001	Representative of the students in the University Commission devoted to the reform of the Pharmacy studies
February 2000 - January 2001	Member of the Pharmaseminar 2001 organizing Committee

LANGUAGE SKILLS

Italian	native speaker
German	Fluent
English	Fluent
French	Fluent

HOBBIES

Skiing (instructor at the Swiss Snow School of Sils i.E.), Tennis, Football, Swimming, Traveling, socializing with people (member of different student societies), Italian Literature, Sailing

REFERENCES

Prof. Dr. B. Ernst
Institute of Molecular Pharmacy
Pharmacenter
University of Basel
Klingelbergstrasse 50
CH-4056 Basel
Switzerland
Tel: +41 61 267 15 50

Prof. Dr. A. Vedani
Biographics Laboratory 3R
Friedensgasse 35
CH-4056 Basel
Switzerland
Tel: +41 61 261 42 56

Prof Dr. T. Langer
Computer Aided Molecular Modeling Group
Institute of Pharmacy
University of Innsbruck
Josef-Moeller-Haus
Innrain 52C
6020 Innsbruck
Austria
Tel: +43 512 507 5252