UNI
BASEL

Computational Modelling of Metal-Mediated Protein-Ligand Interactions

Inauguraldissertation

zur
Erlangung der Würde eines Doktors der Philosophie

vorgelegt der
Philosophisch-Naturwissenschaftlichen Fakultät
der Universität Basel

von
Maurus Hans Schmid
aus Basel (Schweiz)

Basel, Februar 2013

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät auf Antrag von:

Prof. Dr. Thomas R. Ward
Prof. Dr. Markus Meuwly

Basel, den 26. Februar 2013

Prof. Dr. Jörg Schibler
Dekan

creative commons

Namensnennung-Keine kommerzielle Nutzung-Keine Bearbeitung 2.5 Schweiz

---

Sie dürfen:

das Werk vervielfältigen, verbreiten und öffentlich zugänglich machen

Zu den folgenden Bedingungen:

Namensnennung. Sie müssen den Namen des Autors/Rechteinhabers in der von ihm festgelegten Weise nennen (wodurch aber nicht der Eindruck entstehen darf, Sie oder die Nutzung des Werkes durch Sie würden entlohnt).

Keine kommerzielle Nutzung. Dieses Werk darf nicht für kommerzielle Zwecke verwendet werden.

Keine Bearbeitung. Dieses Werk darf nicht bearbeitet oder in anderer Weise verändert werden.

- Im Falle einer Verbreitung müssen Sie anderen die Lizenzbedingungen, unter welche dieses Werk fällt, mitteilen. Am Einfachsten ist es, einen Link auf diese Seite einzubinden.

- Jede der vorgenannten Bedingungen kann aufgehoben werden, sofern Sie die Einwilligung des Rechteinhabers dazu erhalten.

- Diese Lizenz lässt die Urheberpersönlichkeitsrechte unberührt.

Quelle: http://creativecommons.org/licenses/by-nc-nd/2.5/ch/   Datum: 3.4.2009

Acknowledgements

First of all I want to thank my phd supervisors, Thomas Ward and Markus Meuwly, for giving me the opportunity to work on a very exciting and interdisciplinary topic. Being in an experimental and a theoretical group enabled me to learn a lot in diverse fields from theory to practice and to meet many people of various background from physicist to biologist.

I want to thank all my co-workers who shared time with me in the lab (or better office): Especially I want to thank Beat, who, despite being my master-student, is pursuing a career in computational chemistry. Keep it up :-). Further I want to thank everybody from the carbonic anhydrase team: Fabien, Elisa and Tillmann. Thanks as well to the dual anchoring team: Jeremy, Tillmann, Didier, Elisa. I am very grateful to Pierre, Tibor and Marc for revising this thesis, Franziska for her moral support and shared adventures with Valbond. I want to thank everybody I have ever shared the office with (they are too many to mention ;-)), but a special mention deserve Livia, Jarek and Marc, it was always a pleasure with you! The "running team", especially Yvonne and Valentin. Yvonne, I hope you will recover again soon! I compiled a list with everybody I can remember (totally random according to random.org): Tibor, Yonggang, Yvonne, Jing, Christian, Pierre, Jarek, Valentin, Ewa, Franziska, Tristan, Juvenal, Max, Anamitra, Tillmann, Tobias, Marc D., Tommaso, Mark, Thibaud, Ana, Sabina, Prashant, Florent, Marek, Stephan, Frida, Michael, Nuria, Elisa, Raphael, Jeremy, Myung Won, Fabien, Livia, Sarah, Alessia, Didier, Praneeth, Cheikh, Julian, Maxime, Lixian, Anja, Vijay,

Acknowledgements

Marc C, Mishra, Franziska. I am sorry, if I forgot somebody!

Finally I want to thank my family for their support through all this time, and of course 丽娴, 如果没有你，它不会是可能的！ (I hope this makes any sense and that I can write it without google translate one day;-))

Abstract

Although metalloproteins account for nearly half of all proteins in nature, computational modelling of metal-mediated protein-ligand interactions is understudied and molecular mechanics programs and force field parameters compatible to proteins and transition metals are not readily available. Within this thesis, various computational approaches were pursued towards the design of artificial metalloenzymes and the modelling of metal-mediated protein ligand interactions. Four challenges were identified and addressed. The first consisted of the identification of suitable protein scaffolds for the creation of artificial facial-triad motifs. The second part focused on the development of reliable force field parameters for the anchoring of sulfonamide bearing anchors within human carbonic anhydrase 2. In order to reliably predict the geometry of catalytically relevant piano stool artificial cofactors within host proteins, the third part aimed at developing force-field parameters (using the Valbond formalism) for d^6-piano stool complexes. Finally, the fourth step combined the above developments towards the reliable prediction of first and second coordination sphere environments around artificial cofactors/inhibitors.

Contents

Contents

Part I.

Introduction

1. Background

Although metalloproteins account for nearly half of all proteins in nature, computational modelling of metal-mediated protein-ligand interactions is understudied and molecular mechanics programs and force field parameters compatible to proteins and transition metal are not readily available. In this chapter, chemical and biological systems where metal-mediated protein-ligand interactions play an important role are summarized, metal complexes, proteins and metalloproteins are introduced and artificial metalloenzymes based on protein scaffolds are presented.

1.1. Metal Complexes

Since prehistoric times, people have been fascinated by colourful dyes or pigments. Many transition metal complexes have bright colors and naturally occurring pigments such as ochres and iron oxides have been used as colorants. Other complexes have been synthesized since the beginning of chemistry, e.g. Prussian blue, $Fe_7(CN)_{18}$, was produced since the early eighteenth century,[1] but the nature of these compounds remained mysterious. A major breakthrough was made when Alfred Werner proposed that the product of the reaction between of cobalt chloride and ammonia is an octahedral complex, $Co(NH_3)_6]^{2+}$, with dissociated chloride ions.[2] He subsequently developed the basis of modern coordination chemistry and was awarded the Nobel Prize in 1913.

A more modern definition of metal complexes states: "In a coordination complex, a central atom or ion is coordinated by one or more [...] ligands which act as Lewis

bases, forming coordinate bonds with the central atom [...] [that] acts as a Lewis acid." [3] Depending on ligand and central atom, different geometries can result. Common are two- to six-coordinated complexes, with linear (two-coordinate), trigonal planar (three-coordinate), tetrahedral or square planar (four-coordinate), trigonal bipyramidal or square pyramidal (five-coordinate) and octahedral (six-coordinate). Many observed geometries are regular and symmetrical, but electronic effects like the Jahn-Teller distortion or the *trans* influence can lead to deviations from an ideal structure. [4,5] Ligands that coordinate with more than one donor atom are referred to as chelating ligands. Coordination is not limited to single atoms, since ligands can coordinate to a metal center via an uninterrupted series of atoms. This hapticity is indicated by η^n where n indicates the number of atoms that coordinate to the metal. Prominent η-bonding moieties are η^2 with e.g. H_2, η^3 with $H_2C=CH-CHR$ type allyls, η^5 for Cp and Cp^* or η^6 for other arenes. One focus of this thesis are d^6 three-legged piano stool complexes like the $(\eta^n\text{-arene})ML_3$ (M = Ru, Rh, Ir; $n = 5, 6$) shown in Figure 1.1

1.2. Proteins and Metalloproteins

There are three kinds of polymers that are prevalent in nature: nucleic acids, polysaccharides and proteins. Proteins perform a vast array of biological functions within living organisms. They can act as building blocks, transporter or catalysts. A key point for the understanding of the function of a protein is its structure. Proteins and peptides are polymers of amino acids connected via amide linkages. The sequence of amino acid residues is often referred to as primary structure. Apart from this primary ordering, proteins are ordered by elements of secondary structure like helices, β sheets or hairpins. The secondary structure can be formally defined by the patterns of hydrogen bonds between the backbone amino and carboxyl groups, or from the pattern of the backbone dihedral angles of the Ramachandran plot. [7] Assignment is usually done automatically using algorithms like STRIDE [8] that are implemented in many programs. The tertiary
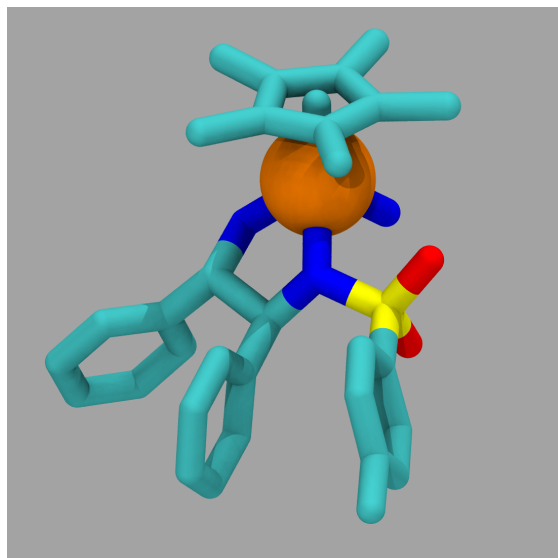
Figure 1.1.: The structure of a typical η^5-piano stool d^6 complex. The iridium is coordinated by an amido-amine chelate, a *Cp**-moiety and an ammonia group. Hydrogen atoms are hidden, the other atoms are coloured by element (carbons in cyan, nitrogens in blue, oxygens in red, sulfur in yellow and iridium as a van der Waals (vdW) sphere in orange). This particular iridium complex catalyses the hydrogenation of ketones.[6]

structure is the folding of the secondary structure and finally the quaternary structure is defined as the association of several protein chains, yielding an oligomeric complex structure.[9] In Figure 1.2 the structure of a typical protein is demonstrated. If the protein can be crystallized, its structure can be solved by X-ray structure analysis. Proteins are flexible rather than completely rigid as highlighted by crystal structures.[10] If a protein is not too large, NMR studies can yield an ensemble of structures or molecular dynamics can simulate the movement.[11] The function of the protein is often closely related on its ability to bind molecules. Many proteins can specifically bind other molecules, be it small molecules, other proteins or nucleic acids.[12,13] Protein-ligand interactions are thus a key point for the understanding of the function and in the center of many studies.

Through recombinant protein expression, for example *E. coli*, it is possible to harvest significant amount of protein for for further experiments. Moreover, biochemical
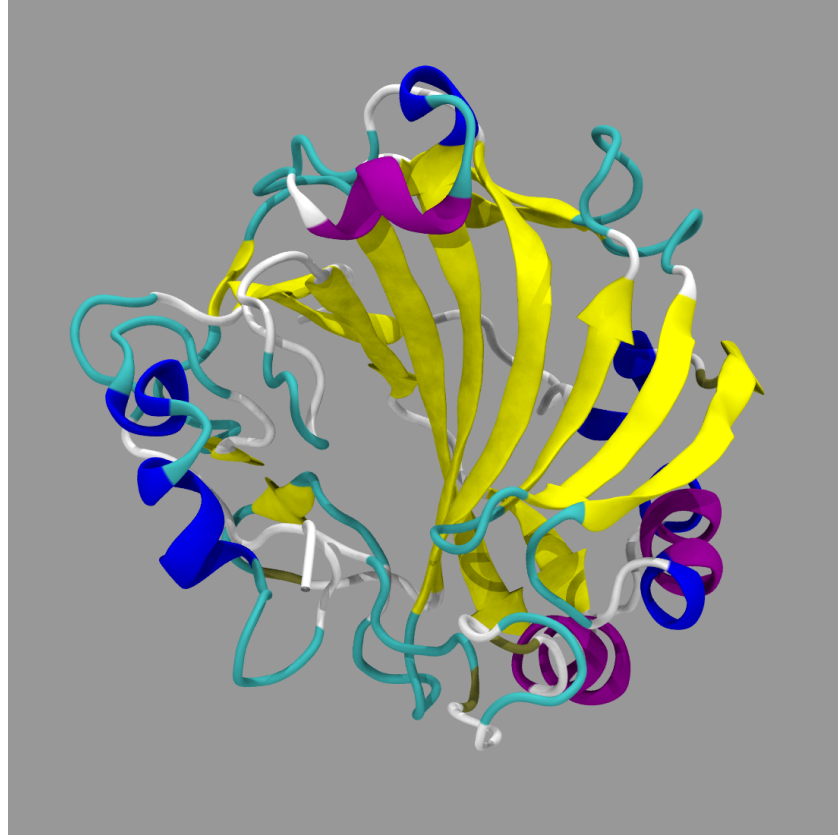
Figure 1.2.: Human carbonic anhydrase 2 features many different secondary structure elements. β-sheets (yellow) are most prominently visible in the center of the protein. A cavity is formed by the β-sheet plane together with some less ordered areas (turns are cyan and random coils are white). α-helices (purple) and 3_{10}-helices (blue) yield additional stability.

techniques such as the polymerase chain reaction[14] or site-directed mutagenesis[15] open up the possibility for manipulation of structure and function.[16]

Metalloproteins The naturally occurring amino acid residues account only for a small proportion of the possible chemical functionality. The diversity of function is greatly enhanced by cofactors such as small organic molecules, single metal atoms or clusters that contain metal and non-metal atoms.[17] Although life is traditionally regarded as organic, there are at least 13 metals which are essential for plants and animals. Only four of them (Na, K, Mg and Ca) are present in large quantities. The other nine (Va,

Cr, Md, Mn, Fe, Co, Ni, Cu, Zn) are present only in small amounts, but imbalances are often harmful and may even be lethal.[18] Most of these trace metals are found as components of proteins, in metalloproteins.[19] Metalloproteins account for nearly half of all proteins in nature.[20] Amino acids that regularly act as metal binders in proteins are thiolate of cysteine, imidazole of histidine, carboxylate (e.g. glutamic and aspartic acid), and phenolate of tyrosine. Depending on metal and its oxidation state, different coordination numbers and geometries can occur.[21]

1.3. Chirality

A molecule is chiral, when its mirror image cannot be superposed on the original molecule. There are different sources of chirality, but the most simple and common case is a tetrahedral atom bearing four different substituents which yields a stereocenter. Chirality is a very intriguing and widespread property in nature: all amino acids but glycine are chiral. Naturally occurring in proteins are almost exclusively L- respectively (S)-amino acids, only in exotic and rare cases (like in the venom of a marine cone snail) D- respectively (R)-amino acids have been found.[22] Proteins are thus chiral polymers and opposite enantiomers of a small molecule can have very different biological effects toward enantiopure proteins. Thalidomide may well be the most famous example of two enantiomers of a drug having adverse effects: one of the enantiomers is a powerful sedative but the other one causes birth defects.[23] Chirality is not limited to carbon atoms nor to organic molecules, because many coordination compounds are chiral too. Figure 1.3 demonstrates chirality on a d^6-piano stool complex. To synthesize a molecule in enantiopure form, there must be a source of chirality present to induce this selectivity. In asymmetric transition metal catalysis, catalysts feature asymmetric groups to induce it. Limitations in the enantioselectivity of the catalyst can be circumvented by step-wise enrichment of one enantiomer, for example through selective recrystallization. However, as this may be tedious and adds additional steps to the synthesis, highly selective catalysts
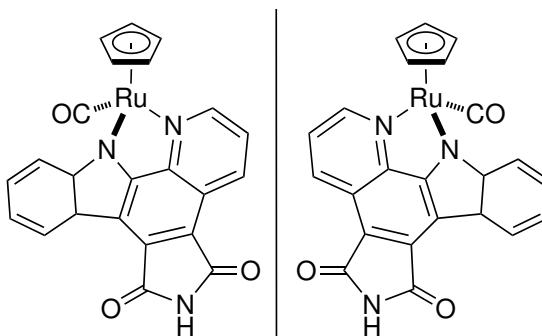
Figure 1.3.: A metal as a chiral center: The two enantiomers of a ruthenium metal complex: The (R) enantiomer on the left is the mirror image of the (S) enantiomer on the right.

are desired.

1.4. Catalysis

One of the most prominent features in chemistry is the concept of catalysis: a substance, that can accelerate a process without being consumed, has a very remarkable characteristic. A catalyst accelerates a reaction by lowering its activation energy, which is achieved by providing an alternative reaction mechanism involving different transition states of lower energy. A reaction that proceeds via a large activation barrier thus may be made accessible in the presence of a suitable catalyst. Jöns Jacob Berzelius was the first to use the term "catalyst" as early as 1836.[24] Catalysts since then have found numerous applications in modern society and in industrial processes. A lot of large scale chemical processes are catalytic processes. One of the most prominent examples is the Haber-Bosch process, where nitrogen gas and hydrogen gas react to ammonia.

$$\text{N}_2 + 3\,\text{H}_2 \rightleftharpoons 2\,\text{NH}_3 \tag{1.1}$$

The artificial supply of nitrogen as a fertilizer was arguably the most important invention of the 20th century and without it "nearly half the world would go hungry" because

natural biofixation of nitrogen is estimated to be only about half of the global need.[25] Without catalyst, this reaction would not be accessible.

Catalysts are commonly divided into three main classes: heterogeneous and homogeneous catalysts and enzymes. The first two are defined depending on whether the catalyst is in the same phase as its substrate. Because of their unique properties, proteins that act as catalysts are termed enzymes and form a class of their own.

1.4.1. Heterogeneous Catalysts

Heterogeneous catalysts are in a different phase than the reactants and products. Most heterogeneous catalysts are solids that work with substrates in solution or gaseous phase. Due to the phase separation, reactions can only occur at the interphases where catalyst and reactants meet. Many industrial processes are performed using heterogeneous catalysts as they are easier to handle than homogeneous catalysts: the substrate and the catalyst are in a different phase, which simplifies workup and separation. Heterogeneous catalysts are applied in a way that maximizes its surface area, to maximize contact with the reactants. Typically, a powder or a high surface material coated with the catalyst is applied. A prominent example is the catalytic converter in a motor vehicle which converts toxic byproducts of the combustion process into less poisonous substances. Most modern catalytic converters are "three way" converters which react carbon monoxide, unburned hydrocarbons and oxides of nitrogen to carbon dioxide, nitrogen and water. The catalytic metals (mostly platinum, but also some other transition metals) are dispersed in a thin layer on a ceramic core to maximize the accessible catalyst surface. Also in the aforementioned Haber-Bosch process a heterogeneous catalyst is involved, in which iron serves as catalyst for the production of ammonia from dinitrogen and dihydrogen.

1.4.2. Homogeneous Catalysts

Homogeneous catalysts function in the same phase as the reactants. Typically the catalysts and the substrates are dissolved in a solvent. Homogeneous catalysts are more efficient than heterogeneous catalysts as all the catalyst molecules are accessible to the substrate. Many of homogeneous catalysts based on organometallic compounds have been developed, featuring most metals in the periodic table. Prominent examples include hydroformylation, hydrogenation, transfer hydrogenation or methathesis. [26–28] A lot of these metal complex catalysts are only soluble in organic solvents though, which prohibits biological compatibility and causes environmental issues. Metal complex catalysts typically have a broad range of substrates and complexes featuring asymmetric ligands can induce stereoselectivety on the substrate.

In recent years, several Nobel Prizes have been awarded to catalysis: in 2001, [29,30] 2005 [31–33] and 2010, [34,35] thus highlighting the importance of catalysis to modern society.

1.4.3. Enzymes

Proteins can exhibit catalytic activities as well, they are commonly considered as the third class of catalysts. Such enzymes are highly active and selective. These enzymes can greatly accelerate the rate of a reaction with a very high specificity. [36] Almost half of all enzymes require the presence of a metal atom to function. [17] Metalloenzymes are metal-containing biopolymers that can catalyse a wide range of important biological processes. Like metalloproteins, metalloenzymes can either incorporate a cofactor binding the metal or incorporate the metal directly bound to amino acid side chains. A well known example incorporating iron with a *Heme*-cofactor is the cytochrome P450 superfamily. Residues featuring free valence electrons like histidine, methionine or carboxylate residues are commonly involved in direct metal binding and special metal binding motifs like the two-histidines one-carboxylate triad have evolved.

The activity of an enzyme can be decreased by inhibitors, usually small (ligand)

molecules that bind to the active site. Enzyme inhibitors occur naturally and are involved in the regulation of metabolism thus many drugs are enzyme inhibitors and other inhibitors are used as herbicides and pesticides. Most inhibitors bind reversible and the inhibition is a result of equilibriums between free, inhibitor-bound and substrate-bound enzyme. Others bind irreversible which involves usually a covalent bond to the protein. Many inhibitors are small organic molecules, but transition metal complexes can be an interesting alternative to purely organic compounds, as they can adopt conformations (e.g octahedral complexes) not accessible to traditional organic molecules and yield high inhibition.[37–39]

Carbonic anhydrase as an example enzyme is discussed in chapter 4.

1.5. Artificial Metalloenzymes

Traditional metal complex catalysts are very versatile but usually not highly specific. Enzymes are highly active and selective, but they do not offer the vast range of catalyzed reaction when compared to metal complexes. The general idea of artificial metalloenzymes is to combine the strengths of these two approaches (i.e. homogeneous and enzymatic catalysis) to generate new, efficient hybrid catalysts displaying the (enantio-)selectivity of enzymes but catalyzing a broad range of reactions using an artificially incorporated metal.

There are several strategies to incorporate a metal into a biopolymer scaffold. The most challenging form of metalloprotein design is *de novo* design, or design 'from scratch'. To bind the metal, a polypeptide sequence is constructed that folds into a defined structure to host the metal ion. As α helical bundles are a common scaffold for a number of heme proteins in nature, heme centers are some of the most common metal centers incorporated into *de novo* α helical bundles.[40] Designing a metalloprotein using *de novo* scaffolds could have advantages like complete control over the protein structure and the metal binding environment, but current knowledge of protein folding limits *de novo*

designed scaffolds to only a few types, such as the α helical bundles.[20]

On the other hand, natural proteins offer a very large variety of folds and usually are quite stable to mutations. So instead of designing the whole system from scratch, existing protein can be used as a scaffold and host for the metal. It is not even necessary to make mutations on the binding site, replacing the natural metal with a different metal and thus exploiting the promiscuity in metal binding of enzymes can yield an enzyme that is more active than its natural form[41] or has a different activity.[42,43] In chapter 3 the availability of promiscuous metal binding sites in proteins is explored.

An existing metal binding site can be redesigned to modify the function or introduce a new metal as demonstrated in the conversion of the Gly II enzyme from a glyoxalase, which hydrolyses thioester bonds, to a β-lactamase.[44] This was achieved by incorporation and adjustment of functional elements through insertion, deletion and substitution of several active site loops, followed by point mutations.[44] Another example is the modification of a copper binding site into a binuclear Cu center by replacing the copper-binding loop of blue copper azurin with a slightly larger Cu-binding loop to yield a bimetallic, purple hybrid construct (see Figure 1.4).[45] A more exotic example is the engineering of an uranyl-specific binding protein from NikR.[46] Although not catalytic, this remarkable example shows the power of protein engineering.

In a different approach, the metal is introduced using an artificial cofactor. Pioneering work was performed by Wilson and Whitesides in the late 1970s when they incorporated a biotinylated rhodium complex into avidin to yield a modest enantioselecivity (44% *ee*) for the reduction of α-acetamidoacrylic acid to *N*-acetylalanine.[47] Since then, artificial metalloenzymes for enantioselective catalysis have gained significant momentum.[48]

The metal can be incorporated using an anchoring group with a strong affinity for a given host protein, e.g. a strong inhibitor or even covalent linking.[49] Figure 1.5 presents a general strategy for the creation of artificial metalloenzymes using a linker to incorporate the metal complex. Systems based on cofactors can be optimized and tuned on several
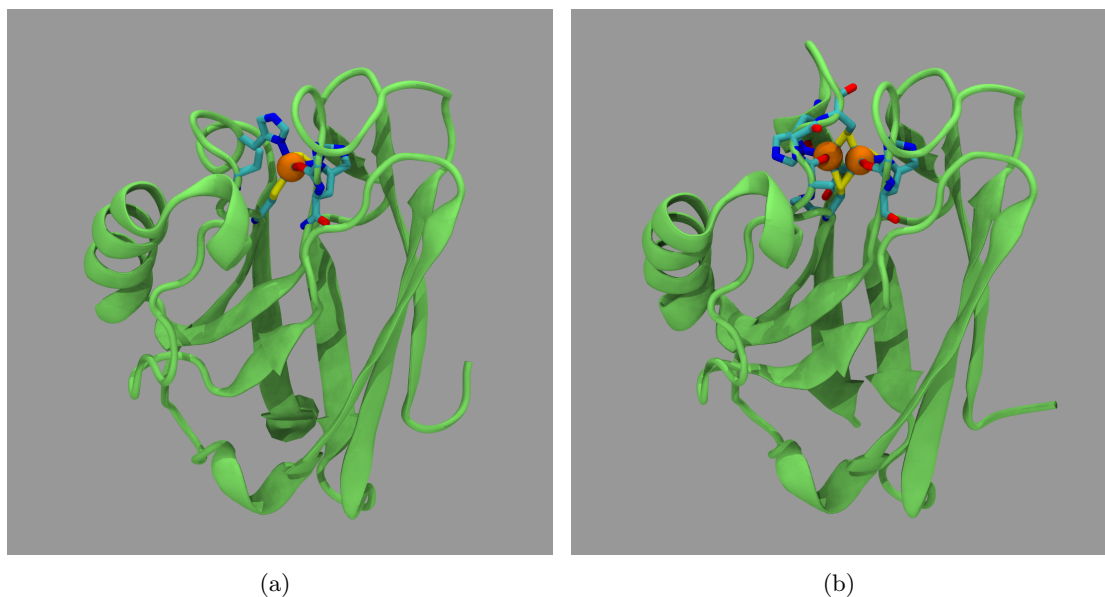
Figure 1.4.: Incorporation of an additional metal through mutation of an existing binding site. (a) Native azurin (PDB entry 4AZU), (b) Bimetallic mutant. The protein (cartoon, green) was mutated by introducing a larger loop featuring an additional methionine. The additional sulphur (yellow) enables bridge between the Cu (orange spheres).

levels: linkers and metal-ligands can be varied and changed, but also mutations on the host protein can influence the reactivity or even invert the selectivity. The spacer length directly sets the metal position. If it is too short, the metal might be buried and not accessible to the substrate or the complex may not fit into the binding pocket at all. If it is too long the metal is too far away from the protein and the influence of the protein is reduced. Mutations of the protein residue in direct vicinity of the bound metal complex influence the reaction by changing the binding mode of the metal complex or operating on the substrate directly by providing a different environment for the reaction. All these interactions can influence each other and design may not be straight-forward. Computational tools can give valuable insights and help in understanding and developing the catalyst further.

Hybrid catalysts are not limited to artificial metalloenzymes. Besides using proteins
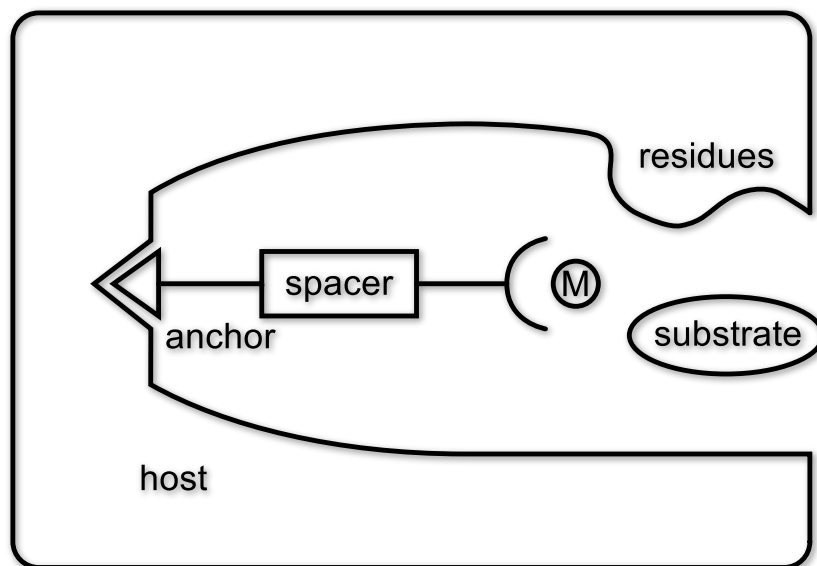
Figure 1.5.: Scheme of an artificial metalloenzyme using a linker approach. The metal (M) is bound to the scaffold (host) with an anchor. A spacer can be used to fine-tune the metal position. The protein residues in the vicinity of the metal influence the reaction by influencing the metal and the substrate.

as scaffold for the metal complex, other supramolecular entities have been envisaged. For example hybrid catalysts involing DNA,[50,51] peptides[52,53] and antibodies[54] have been introduced.

1.5.1. Artificial Metalloenzymes based on the Biotin Streptavidin Technology

The Ward group has extensive experience with artificial enzymes based on the biotin streptavidin technology where the biotin is used as a linker to incorporate the transition metal complex into the protein. Hybrid catalysts for hydrogenation,[55,56] transfer hydrogenation,[57–59], sulfoxidation[60], allylic alkylation,[61] methathesis,[62] dihydroxylation of olefins[63] and C-H activation[64] have been developed.

Streptavidin (Sav) has an extraordinarily high binding affinity to biotin. It is a homo-tetramer. All four of the monomers can bind a biotin thus one protein can incorporate

four ligands. The binding sites for two of these monomers face each other, forming a large vestibule on either side of the tetramer (see Figure 1.6). The high affinity for biotin ($K_d \approx 10^{-14}$ mol/l) is used to anchor a metal complex to the protein. Provided that the metal complex is not too big, one complex can fit in each of the four binding sites. Mutations of the residues in the vestibule have the highest influence on catalysis, influencing conversion and enantiomeric excess significantly.

(a)   (b)
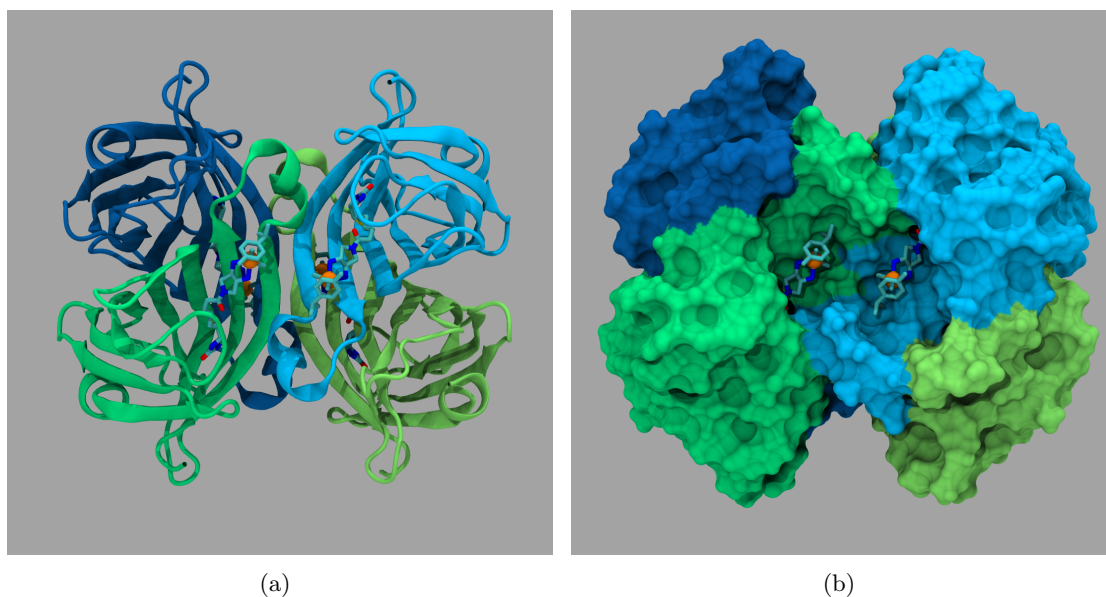
Figure 1.6.: (a) Cartoon representation of a streptavidin based artificial metalloenzyme (PDB code 2WPU) catalyzing transfer hydrogenation.[57] The four Sav subunits are coloured in shades of blue and green. The biotinylated metal complex (stick) features an ruthenium (orange sphere) bound to two nitrogens (blue). Carbons are cyan, oxygens red and sulphurs are yellow. (b) Solvent accessible surface: the wide cavity featuring the metal complexes is in the center of the protein and is constructed from two monomers.

2. Computational Methods

The basis for computational chemistry was set out much earlier than computers actually existed. Theories like valence bond theory or molecular orbital theory had been around much longer and when Schrödinger, published his first paper on quantum mechanics in 1926, computers had not been invented until thirty years later. With modern computers, application of these theories became accessible and as technology advances, calculations can be performed on increasingly large systems with more and more accuracy.

Computational methods can be broadly divided into two large groups, quantum mechanical (QM) and molecular mechanical (MM) methods. Whereas the former are based on quantum chemistry, the latter are highly parametrized and use more simple mathematical functions that describe the potential energy of the system.[65,66]

First, the quantum mechanical approach is outlined, then a more detailed introduction into force fields especially concerning metals is given. Finally valence bond theory and Valbond are presented.

2.1. Quantum Mechanical Methods

There are several quantum chemical methods in computational chemistry. Usually the Born-Oppenheimer approximation is assumed and the method is based on the time-independent electronic Schrödinger equation:

$$(\hat{H}_{el} + V_{NN})\psi_{el} = U\psi_{el} \tag{2.1}$$

One group of methods solves the Hartree-Fock (HF) equations, optionally followed by one of a number of post-HF methods for including the effect of the electron correlation. Where the Hamiltonian operator $\hat{H}_{el}$ is purely electronic. Further the variational theorem is applied and electron correlation is neglected. For metal-complexes, due to the high number of electrons involved leading to slow calculations, mainly the Density Functional Theory (DFT) method is used.

2.1.1. Density Functional Theory

The original idea of DFT is to use the electron density with 3 degrees of freedom instead of the N-electron wavefunction with $3N$ degrees of freedom. Hohenberg and Kohn proved with a theorem that the ground state energy and other properties of a system are unambiguously defined by the electron density.[67] That is, the energy is a functional of the electron density. Later, Kohn and Sham suggested a practical way to solve the Hohenberg-Kohn theorem.[68] In this approach, the electron density is expressed as a linear combination of basis functions similar to HF orbitals. The accuracy of the method can be similar to the more costly post-HF methods, however, as it is dependent on $3N$ degrees of freedom again, similar calculation speeds as for HF are achieved.

There are various functionals. Some functionals were developed from quantum mechanics, others were developed by parametrizing functions to reproduce experimental results. The functionals based on the electron density only, are called local density approximation (LDA) or for high-spin systems local spin density approximation (LSDA). Generalized gradient approximations (GGA) also take into account the gradient of the density. Hybrid methods combine functionals with a portion of exact exchange from Hartree-Fock theory, weighed with empirical parameters to give the most accurate results. The most popular functional, B3LYP (**B**ecke, **three**-parameter, **L**ee-**Y**ang-**P**arr[69,70]), is

a hybrid functional with three empirical parameters. It is defined by:

$$E_{xc}^{B3LYP} = (1-a)E_x^{LSDA} + aE_x^{HF} + b\Delta E_x^B + (1-c)E_c^{LSDA} + cE_c^{LYP} \tag{2.2}$$

The exchange correlation energy E_{xc} is mixed of several exchange and correlation formulas, where a,b and c are parameters that are optimised.

2.2. Molecular Mechanics Methods

QM methods are very accurate, but for larger molecules calculations become increasingly difficult due to the bad scalability with system size that results in prohibitive computational costs.

Force fields try to energy landscapes using simple and fast assumptions to reduce the computational effort of calculating physical properties. Most general-purpose force fields for atomistic simulations (e.g. CHARMM[71], Amber[72], Gromos[73] or OPLS[74]) are defined as a sum of bonded and nonbonded energies

$$V = \sum E_{bonded} + \sum E_{nonbonded} \tag{2.3}$$

In standard CHARMM, bonded interactions are considered for atoms that are within three bonds of each other whereas nonbonded interactions are considered for the rest. But other schemes exist, where, e.g., 1-4 interactions are scaled to reduce their effect.

In the general force fields above, the bonded energies are sums over harmonic potentials for stretching and bending terms, periodic functions for torsions. Improper terms are used to enforce planarity.

$$E_{\text{bonded}} = \sum_{\text{bonds}} k_r(r-r_0)^2 + \sum_{\text{angles}} k_\theta(\theta-\theta_0)^2 + \sum_{\text{torsion}} k_\gamma(1+cos(n\omega-\gamma)) + \sum_{\text{improper}} k_\phi(\phi-\phi_0)^2 \tag{2.4}$$

2. Computational Methods

A more realistic description of a covalent bond at higher stretching would be provided by the computationally more expensive Morse potential. Corrections such as Urey-Bradley are commonly included as well. These terms can be further complemented by additional terms describing special interactions like hydrogen bonds[75-77] in more details. Cross-terms that describe nonlinear interactions can improve structures and vibrational forces by e.g., elongating bonds in dependence of an angle. But an accurate determination of cross-term parameters requires extensive reference data.[78,79]

The nonbonded interactions typically include electrostatic interactions and van der Waals forces. In most simple force fields, these are represented using Coulomb's law and a Lennard-Jones potential, respectively:

$$\sum E_{nonbonded} = \sum_{i<j} \left\{ 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] + \frac{q_i q_j e^2}{r_{ij}} \right\} \qquad (2.5)$$

where the double sum is over all pairs of atoms that are not bonded to one another or to a common atom. The standard 6-12 repulsion-dispersion (Lennard-Jones) potential can be modified with other exponents to yield different energy surfaces.[80] Technically, for every pair of atoms, the parameters ϵ_{ij} and σ_{ij} would need to be fitted. To avoid this, mixing rules for these parameters from parameters per element have been developed. In CHARMM, the Lorentz Berthelot rules[81,82] are applied, but other combination rules like geometric means can be used.[80] Instead of the simple point charges, multipoles may be implemented to cover inhomogeneous charge distribution.[83,84] Polarizability describes the tendency of a charge distribution to be distorted from its normal shape by an external influence like another atom or molecule. In most force fields it is neglected, but especially for calculation of interaction energies involving metal complexes it may not negligible.[85]

A fundamental assumption of the molecular mechanics method is transferability of the force field. Parameters determined with a set of model compounds, are assumed to be valid for similar molecules too. To ensure transferability, atoms with similar

environment are grouped to a common *atom type* with common parameters. These atom types determine what parameters apply for all interactions. The performance of a force field depends on the functional form of the energy and the quality of the parameter set as well as on the atom type definitions. The more atom types are introduced, the more exact a force field can become, as more parameters are introduced. However, this also increases the effort for parametrization.

The nonbonded terms are most computationally intensive because they include many more interactions per atom. The usual approach to limit the computational effort includes cutoffs and thus neglecting nonbonded interactions beyond a certain distance. Although this approach introduces non-continuity, a cutoff large enough and suitable switching or shifting functions can reduce the errors.[86] Alternatively, for long-range electrostatic interactions in periodic systems, Ewald summation[87] or its implementation as Particle Mesh Ewald (PME[88]) method can be used.

A drawback of common force fields is the fixed topology: bonds cannot be broken or formed during the simulation.

2.2.1. Metals in Force Fields

Already in the beginnings of force field calculations, computations on inorganic molecules were performed. For example, Corey and Bailar in 1959 examined the stereochemistry of diverse transition metal complexes.[89] But simulation of inorganic compounds never became as prevalent and routine as simulation of purely organic molecules. Although some inorganic systems are not much different to organic systems from a force field point of view, a lot of metal complexes have properties that are difficult to model using the standard terms.[66,90–94] They have a much wider range of geometries involving not only one-, two-, three- and four-bonded structures (which are typical for organic molecules), but also five-, six or even higher coordination numbers. Depending on their coordination number, complexes can form square planar or tetrahedral shapes (four-coordinate),

square pyramidal or trigonal bipyramidal shapes (five-coordinate) or octahedral shapes (six-coordinate). Force fields typically are not able to cope with such structures of high symmetry. For example in octahedrons, atoms can form angles of either $90°$ or $180°$ with the same atom types. One needs a functional form featuring several minima or separate atom types for all of the atoms, resulting in an extensive number of parameters. Also, non-equilibrium angles far from the minimum or equilibrium around $180°$ are problematic for the harmonic approximation applied in most general force field and yield unphysical results. Additionally, the ideal structures are often heavily distorted from the ideal shape due to electronic effects like the Jahn Teller effect or *trans* influence.[4,5,95] The harmonic potential is not suitable to model these large distortions and special functional forms have been developed. The Universal Force Field (UFF[96]) for example, uses a cosine Fourier series for each angle:

$$V = k_\theta \sum_{n=0}^{m} C_n \cos n\theta \qquad (2.6)$$

Coefficient C_n adjusts the minimum angle whereas m is adapted to get the number of minima required for this angle. What makes things even more complicated are binding modes not present in organic molecules: especially η-type binding poses a challenge, as the classical bond and angle model of force fields does not apply.

Several approaches for the treatment of coordinative bonds can be imagined[65,97,98]: *Bond free coordination:* only nonbonded interactions are included in the metal-ligand interactions and all forces stem from vdW and electrostatics contributions. This also has the advantage of allowing ligand exchange during simulation. One example for this approach would be the zinc parametrization used as default in CHARMM.[99] However, no orientation constraints are explicitly set which can result in unwanted structures and coordination numbers.[100] *Angle free model:* all metal-ligand bonds are included, but all metal-centered angles are ignored. *Covalent model:* the metal is treated the same way as all other atoms. But using a full bonding model with standard potential energy functions can introduce the problems described above. Although bonded interactions preserve the

observed geometry of the metal-binding they are limited to fixed geometries. *Pseudo-atom approaches:* Dummy (pseudo) atoms can be introduced to help modelling challenging structures. *Specialized Force Fields:* To circumvent all these problems, specialized force fields can be used.

2.3. Specialized Metal Force Fields

Apart from using a classical force field setup and treating the metal as any other atom, very diverse approaches have been taken on how to calculate metal containing systems. Some of these specialized methods make additions on top of normal force fields, but also completely unrelated approaches with little similarity to classical force fields have been developed. Some of the more important and relevant methods are briefly summarized below.

2.3.1. YETI

YETI[77,101] is a program with force field for the calculation of metal interactions. It uses conventional point charge electrostatics and Lennard-Jones vdW terms from AMBER[72], but adds a directional function for hydrogen bonds and salt linkage. Metal centers are treated with a special function with parameters depending on coordination pattern and involved atom types:

$$
\begin{aligned}
E_{\mathrm{MC}} = \sum_{\mathrm{ML\ pairs}} \left(\frac{A''}{r_{\mathrm{M\cdots L}}^{12}} - \frac{C''}{r_{\mathrm{M\cdots L}}^{10}} \right) + (E_{\mathrm{MC}} + E_{\mathrm{LFS}}) \cdot \\
\prod_{\substack{\mathrm{independent} \\ \mathrm{angles}}} \cos^2 \left(\psi_{\mathrm{L\cdots M\cdots L}} - \psi_0 \right) \cdot \frac{1}{n} \sum_{\substack{n \\ 1^{st}\ \mathrm{shell} \\ \mathrm{ligands}}} cos^n \left(\omega_{\mathrm{M\cdots L-LP}} \right)
\end{aligned}
\tag{2.7}
$$

YETI uses a 10/12 type Lennard Jones rather than a harmonic function for the metal-1ligand distance whereas the angle-dependent energy has parameters for the complex type, E_{MC} and the ligand-field stabilization energy, E_{LFS}. A further parameter is ψ_0 which is

extracted from experimental structures or set according to symmetry desired. Finally a term depending on the ligands of the first coordination sphere ($\sum cos^n\left(\omega_{\text{M}\cdots\text{L}-\text{LP}}\right)$) is considered. The metal-center function has been calibrated for Zn(II), Co(II), Cu(II), Ca(II), Mg(II), Ni(II), Fe(II), and Fe(III).[a] It allows for the simulation of tetrahedral, square-planar, square-pyramidal, trigonal-bipyramidal and octahedral coordination.

Yeti has mainly been used (in conjunction with other programs) for determining the toxicity of small molecules towards selected proteins.[102]

2.3.2. Ligand-Field Molecular Mechanics

The Ligand-Field Molecular Mechanics (LFMM) approach[103–105] extends conventional molecular mechanics by explicitly incorporating the ligand field stabilization energy (LFSE) for d electrons. The observed structure is then a compromise between steric and electrostatic ligand-ligand interaction and the d electronic stabilization. The LFSE term is added on top of a standard force field, typically a MOE implementation[106] in conjunction with the AMBER[72] is used. In this implementation, the LFSE is based on the angular overlap model (AOM).[107] The d orbital energies are derived by diagonalizing the 5×5 ligand field potential matrix, V_{LF}. The matrix elements of V_{LF} are given by equation:

$$\langle d_i|V_{LF}|d_j\rangle = \sum_{l}^{N} \sum_{k}^{symm} F_{ik}^l F_{kj}^l e_k^l \tag{2.8}$$

where d_i and d_j are d orbitals. The sums extend over N ligands and three $symm$ binding modes: One σ, and two π bonds (π_x and π_y, respectively). The F factors depend on the angular coordinates of the ligands and essentially describe the overlaps between the d functions and appropriate ligand orbitals. The AOM energy parameters, e_σ, e_{π_x}, and e_{π_y} are expressed as a function of the bond length (r):

$$e_\gamma = a_0 + a_1 r + a_2 r^{-2} + a_3 r^{-3} + a_4 r^{-4} + a_5 r^{-5} + a_6 r^{-6} \tag{2.9}$$

---

[a]http://www.biograf.ch/index.php?id=software&subid=yeti

where the a_i are empirically determined parameters but in practice only one a_i is usually used.

LFMM has been applied to various systems, ranging from d^9 copper centers[108] to copper proteins[109].

2.3.3. SIBFA Polarizable Force Field

SIBFA[83] (Sum of Interactions Between Fragments Ab initio computed) is a polarizable force field formulated and calibrated on the basis of *ab initio* supermolecule computations.[85] SIBFA uses a fragment based approach where the molecule is divided into suitable rigid blocks. Two successive fragments are connected with bonds, conformational changes take place only by rotations around the bond between the two fragments. It is designed to mimic interaction energies obtained from quantum chemical calculations typically at the MP2 level. The intra-molecular energy (E_{intra}) in a flexible molecule is computed as a sum of inter-fragment interaction energies between the molecule fragments, using a similar formulation as for the intermolecular interaction energy. The SIBFA intermolecular interaction energy (E_{int}) is formulated as a sum of five contributions

$$\Delta E_{\mathrm{int}} = E_{\mathrm{MTP}} + E_{\mathrm{rep}} + E_{\mathrm{pol}} + E_{\mathrm{ct}} + E_{\mathrm{disp}} \tag{2.10}$$

representing the multipolar electrostatic (E_{MTP}), short-range repulsion (E_{rep}), polarization (E_{pol}), charge-transfer (E_{ct}), and dispersion (E_{disp}) contributions. Electrostatic interactions in SIBFA (E_{MTP}) are calculated using multipoles derived from the Hartree-Fock wave function of the constitutive fragments. E_{rep} is formulated as a sum of bond-bond, bond-lone pair, and lone pair-lone pair interactions. For E_{pol} a polarizing field is computed with the same multipoles as for E_{MTP}. Multipoles and polarizabilities can be obtained from *ab initio* calculations performed on a molecule or molecular fragment previously. Each molecular entity is stored in the SIBFA library of fragments and used

for subsequent assembly of molecules or molecular complexes. E_{ct} introduces coupling between polarizations, E_{pol} and E_{ct} components are fitted. E_{disp} is computed as a sum of $1/R^6$, $1/R^8$ and $1/R^{10}$ terms.

The SIBFA force field has been refined multiple times and applied on various systems, ranging from small zinc systems[110,111] or water clusters[112] to peptides[113] proteins[114].

2.4. QM/MM Simulations

QM methods are very accurate and available for a very broad range of systems including metal complexes. But for larger molecules calculations become increasingly difficult due to the bad scalability with system size which results in prohibitive computational costs. On the other hand, force field methods can be very fast, but most force fields are specialized on a class of systems and they would need extensive parametrization for reasonable accuracy.

QM/MM (quantum mechanics/molecular mechanics) tries to combine the advantages of both approaches and divides the system in a QM and a MM part. Where high accuracy is needed, the interaction energies are calculated using an expensive but accurate QM method, the rest of the system is calculated using a fast MM method. The advantages of this mixed approach have found application on many different systems, and there are several approaches and programs capable of QM/MM.[115] ONIOM (our own *n*-layered integrated molecular orbital and molecular mechanics[116]) for example, is a computational approach implementing a multilayered approach for calculating different parts of a system with different accuracy. This method is very versatile and can be used on biomolecular systems as well as transition metal complexes and catalysis. Another method is the SCC-DFTB (Self-Consistent Charge Density-Functional Tight-Binding) method.[117,118] It is based on a approximate density functional theory method and was implemented as a QM/MM method into CHARMM.[119] It is limited to atoms and systems it was parametrized for, excluding most of the transition metals.

But the QM/MM approach has some disadvantages. Artifacts might emerge as the two (or more) regions need to exchange forces and energies. Usually this is done either using additional link atoms, special boundary atoms calculated within both regions or orbital schemes.[115] The accuracy and speed of a QM/MM calculation depend mainly on the QM method invoked and the size of the QM region, as the larger MM part is typically not the bottleneck of the calculation. Even though QM/MM models have been applied to metal-containing proteins[120,121], their computational requirements remain too high for routine usage in the design of metalloenzyme inhibitors, thus molecular mechanics and empirical force fields are widely used for this purpose.

2.5. Valence Bond Theory

Lewis structures are one of the most widely used ways to describe molecules by chemists since Gilbert N. Lewis introduced them in 1916.[122] Although quite simple, they have the power to describe bonding of various chemical systems. Lewis structures can be seen as the foundation of valence bond theory.[123] Valence bond theory was originally developed by Pauling.[124] Although the availability of fast computers has shifted the focus to molecular orbital theory, it is still a powerful tool to describe molecular interactions.[125] In valence bond theory, chemical bonding is explained using the methods of quantum mechanics, it describes bonding with bonding orbitals. The bonding orbitals are formed by the atom orbitals of each of the bonding partners. Orbitals are mathematical functions that describe the electron probability density. As a result, they form distinctive shapes (see Figure 2.1 for some graphical illustrations). Orbitals are organized in electron shells around the atom core, to yield 1, 2, 3, ... subshells which is described by the principal quantum number (n). These subshells are further divided into s, p, d, f, ... orbitals, described by the angular or orbital quantum number l. The magnetic quantum number m_l generates degeneracy for the orbitals and finally the electron spin quantum number m_s designates the direction of the electron spin. The quantum scheme is summarized in

table 2.1. For many applications it is sufficient to only consider valence shells and closed shells can be neglected.

Table 2.1.: Quantum scheme for the first 3 shells. The quantum numbers and their respective orbital are summarized. In valence bond theory the electron spin quantum number is considered by assigning 2 electrons to each of the orbitals.

n	l	m_l	m_s	Orbital
1	0	0	$+1/2, -1/2$	$1s$
2	0	0	$+1/2, -1/2$	$2s$
	1	$-1, 0, 1$	$+1/2, -1/2$	$2p_x,\ 2p_y,\ 2p_z$
3	0	0	$+1/2, -1/2$	$3s$
	1	$-1, 0, 1$	$+1/2, -1/2$	$3p_x,\ 3p_y,\ 3p_z$
	2	$-2, -1, 0, 1, 2$	$+1/2, -1/2$	$3d_{xy},\ 3d_{xz},\ 3d_{yz},\ 3d_{x^2-y^2},\ 3d_{z^2}$
		$\vdots$		

Orbitals only depend on the Bohr atomic radius a_0 and the polar coordinates (r, θ and ϕ). The hydrogen wavefunctions or atomic orbitals are defined the following way:

$$1s: \quad \Psi = \left(\frac{1}{\pi a_o^3}\right)^{1/2} e^{-r/a_0} \tag{2.11}$$

$$2s: \quad \Psi = -\frac{1}{4}\left(\frac{1}{2\pi a_o^3}\right)^{1/2}\left(2 - \frac{r}{a_0}\right)e^{-r/2a_0} \tag{2.12}$$

$$2p: \quad \Psi = \frac{1}{4}\left(\frac{1}{2\pi a_o^5}\right)^{1/2} re^{-r/2a_0} \begin{cases} \cos\theta & (2p_z) \\ \sin\theta\cos\phi & (2p_x) \\ \sin\theta\sin\phi & (2p_y) \end{cases} \tag{2.13}$$

$$3s: \quad \Psi = -\frac{1}{9}\left(\frac{1}{3\pi a_o^3}\right)^{1/2}\left(3 - \frac{2r}{a_0} + \frac{2r^2}{9a_0^2}\right)e^{-r/3a_0} \tag{2.14}$$

$$3p: \quad \Psi = -\frac{1}{27}\left(\frac{1}{2\pi a_o^5}\right)^{1/2}\left(2 - \frac{r}{3a_0}\right)re^{-r/3a_0}\begin{cases} \cos\theta & (3p_z) \\ \sin\theta\cos\phi & (3p_x) \\ \sin\theta\sin\phi & (3p_y) \end{cases} \tag{2.15}$$

$$3d: \quad \Psi = \frac{1}{81}\left(\frac{1}{6\pi a_o^7}\right)^{1/2} r^2 e^{-r/3a_0}\begin{cases} (3\cos^2\theta - 1) & (3d_{z^2}) \\ 2\sqrt{3}\sin\theta\cos\theta\cos\phi & (3d_{zx}) \\ 2\sqrt{3}\sin\theta\cos\theta\sin\phi & (3d_{yz}) \\ \sqrt{3}\sin^2\theta\cos2\phi & (3d_{x^2-y^2}) \\ \sqrt{3}\sin^2\theta\sin2\phi & (3d_{xy}) \end{cases} \tag{2.16}$$

Most relevant for first-row transition metals are the $3s$ orbital, the three $3p$ orbitals ($3p_x$, $3p_y$ and $3p_z$) and the five $3d$ orbitals ($3d_{z^2}$, $3d_{xz}$, $3d_{yz}$, $3d_{xy}$ and $3d_{x^2-y^2}$) of the valence shell.

Bonding Orbitals When two s orbitals interact, they form a σ-bond, whereas two parallel p orbitals form a π-bond. σ-bonds can be formed by either two s orbitals, by a p-orbital interacting with an s orbital or by two p orbitals in direction of the bond axis. d orbitals can form σ, π or δ bonds. Whereas σ-bonds have no nodal planes along the bonding axis, π-bonds have one nodal plane, δ-bonds have two nodal planes and ϕ-bonds have three nodal planes. See Figure 2.2 for graphics of σ and π bonds. ϕ bonds, formed by two f-orbitals have been purported to exist in U_2.[127] U_2 forms a complex binding

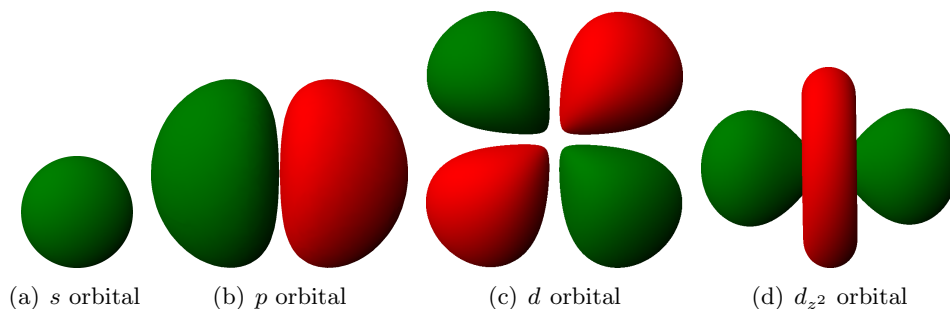(a) s orbital   (b) p orbital   (c) d orbital   (d) d_{z^2} orbital

Figure 2.1.: s (a), p (b) and d (c) type orbitals. The p_x, p_y and p_z respectively d_{xz}, d_{yz}, d_{xy} and $d_{x^2-y^2}$ are identical in shape but not orientation, they are rotation symmetric. The d_{z^2} orbital (d) has a very distinct shape. Positive and negative lobes are green respectively red. These and the following (figures 2.2 and 2.3) graphics are mathematically correct, but not to scale. They were generated using the Orbital Viewer software.[126]

pattern, with with one clear σ and two clear π bonds, further there are several singly occupied δ and finally ϕ bonds. Summed up this yields a quintuple bond.

(a) σ orbital   (b) σ orbital   (c) π orbital
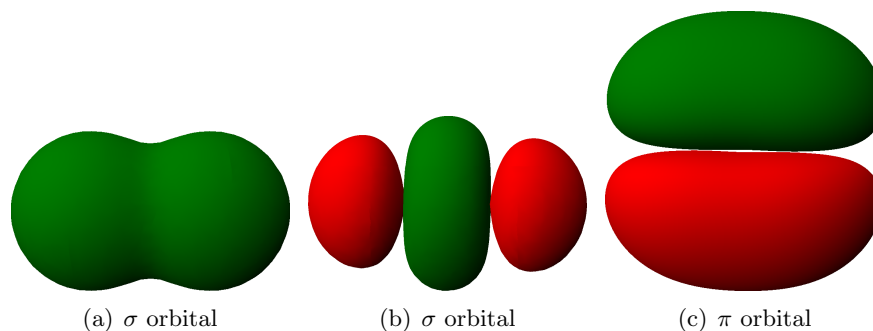
Figure 2.2.: σ and π bonding orbitals. (a) a σ orbital between two s atom orbitals, (b) a σ orbital between two p atom orbitals and (c) a π orbital between two p atom orbitals.

Hybrid Orbitals   For a molecule like CH_4 with four equivalent bonds, the atomic orbital description needs to be changed. Although there are four valence orbitals on the carbon ($2s$, $2p_x$, $2p_y$ and $2p_z$), they are not equivalent. To circumvent this, the concept of hybridization was introduced: the valence orbitals are mixed to form hybrid orbitals. Hybrid-orbitals are linear combinations of s, p, d, f (etc.) orbitals. The CH_4 carbon thus

has a hybridization of sp^3, the eight valence electrons from the s and the three p orbital are distributed in a set of four equivalent sp^3 hybrid orbitals forming the four equivalent bonds arranged in a tetrahedron. The sp^3 orbitals have the following wavefunctions:

$$\psi_1 = \frac{1}{2} \left(2s + 2p_x + 2p_y + 2p_z\right) \tag{2.17}$$

$$\psi_2 = \frac{1}{2} \left(2s - 2p_x + 2p_y - 2p_z\right) \tag{2.18}$$

$$\psi_3 = \frac{1}{2} \left(2s + 2p_x - 2p_y - 2p_z\right) \tag{2.19}$$

$$\psi_4 = \frac{1}{2} \left(2s - 2p_x - 2p_y + 2p_z\right) \tag{2.20}$$

Similarly, sp^2 hybrids form a trigonal planar as observed in BF_3 and sp hybrids a linear structure as in CO_2. In Figure 2.3 the sp, sp^2 and sp^3 orbitals are depicted.

(a) sp orbital   (b) sp^2 orbital   (c) sp^3 orbital   (d) four sp^3 orbitals
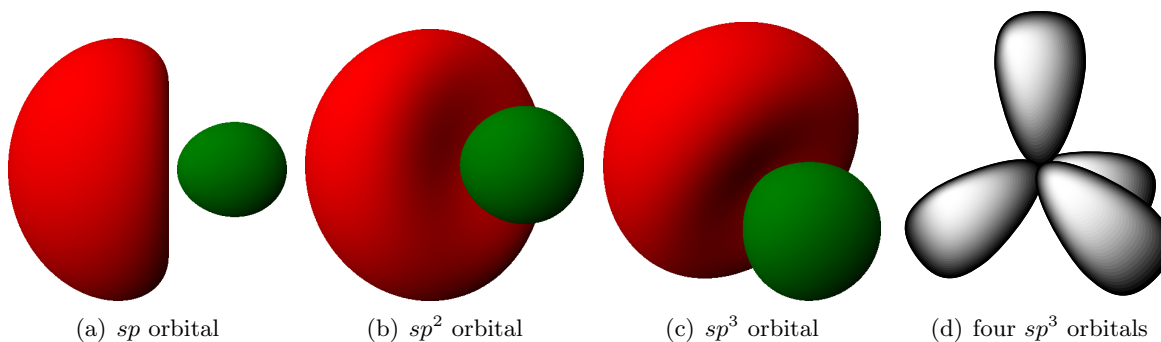
Figure 2.3.: sp (a), sp^2 (b) and sp^3 (c) orbitals. In CH_4 four sp^3 orbitals are ideally suited to describe the tetrahedral structure (d). Whereas Figures (a) to (c) are mathematically correct, Figure (d) is an illustrative scheme commonly used to represent sp^3 hybridization with four sp^3 orbitals.

2.5.1. Valbond

Pauling established fundamental rules to describe covalent bonds [124]:

"1. The electron-pair pond is formed through the interaction of an unpaired electron on each of two atoms. [...]

5. [...] for a given eigenfunction, the bond will tend to be formed in the direction with the largest value of the eigenfunction. [...]"

These concepts can be used to derive a potential that describes the energy as a function of a bond-angle which determines the overlap of orbitals involved.

Valbond [128–131] is based on valence bond theory. It replaces the conventional harmonic bending term in a general-purpose force field with the VB-energy expression which is based on hybrid orbital strength functions as the basis for a molecular mechanics expression. These functions not only describe the energy of bond angles around the minimum, but also at larger distortions where the harmonic approximation breaks down. In this fashion, VB can reproduce unusual geometries as illustrated by the distorted trigonal prism reported for $[W(CH_3)_6]$. [132] More recently, we extended the Valbond formalism to include the *trans*-influence which yields Valbond-trans (VBT). [95]

The derivation of a general expression for the strength of hybrid orbitals follows that of Pauling [133]. For an angle α between non-hypervalent bonds with an $sp^m d^n$ hybrid orbital, two hybrid orbitals are given by the following expressions, respectively:

$$\psi_{\text{hy}_1} = \sqrt{\frac{1}{1+m+n}} (1s + \sqrt{m} \cdot p_z + \sqrt{n} \cdot d_{z^2}) \tag{2.21}$$

$$\psi_{\text{hy}_2} = \sqrt{\frac{1}{1+m+n}} \left(1s + \sqrt{m}(\cos\alpha \cdot p_z + \sin\alpha \cdot p_x) + \right.$$
$$\left. \sqrt{n} \left(\frac{1}{2}(3\cos^2\alpha - 1) \cdot d_{z^2} + \frac{\sqrt{3}}{2}\sin^2\alpha \cdot d_{x^2-y^2} + \sqrt{3}\sin\alpha\cos\alpha \cdot d_{xz} \right) \right) \tag{2.22}$$

The overlap between the two hybrid orbitals orbitals, Δ, is expressed as a function of the hybridizations m and n and the angle:

$$\Delta = \frac{1}{1+m+n}(1 + m\cos\alpha + \frac{n}{2}(3\cos^2\alpha - 1)) \tag{2.23}$$

The overlap is used to define the strength function S:

$$S(\alpha) = S^{max} \sqrt{1 - \frac{1 - \sqrt{1 - \Delta^2}}{2}} \tag{2.24}$$

Where S^{max} is the maximum of the strength function between the orbitals.

$$S^{max} = \sqrt{\frac{1}{1 + m + n}}(1 + \sqrt{3m} + \sqrt{5n}) \tag{2.25}$$

The energy contribution of one (bond-)orbital to the angular energy is:

$$E_{nonhyp} = k(S^{max} - S(\alpha)) \tag{2.26}$$

The parameter k is a scaling constant for the interactions between the atom at the center of the angle with one of its bonding partners. Finally the total energy of this angle is the sum of the two orbitals of the bonding partners. The resulting energy functions have one or two minima at specific angles defined by the hybridization (see Figure 2.4). For example, sp^3 hybridization yields a minimum at 109.5° whereas sp^3d^2 hybridization yields two minima at 90° or 180°, respectively.

Valbond also supports hypervalent compounds using a 3-center 4-electron (3c4e) bonding model.[129,130] For transition metals, Valbond traditionally considers only sd hybrids. Thus complex centres that count more than 12 electrons in their valence orbitals are considered hypervalent and 3-center-4-electron bonds are used instead of the p-orbitals.

The following function is used for the bending energy of the hypervalent 3c4e bonds:

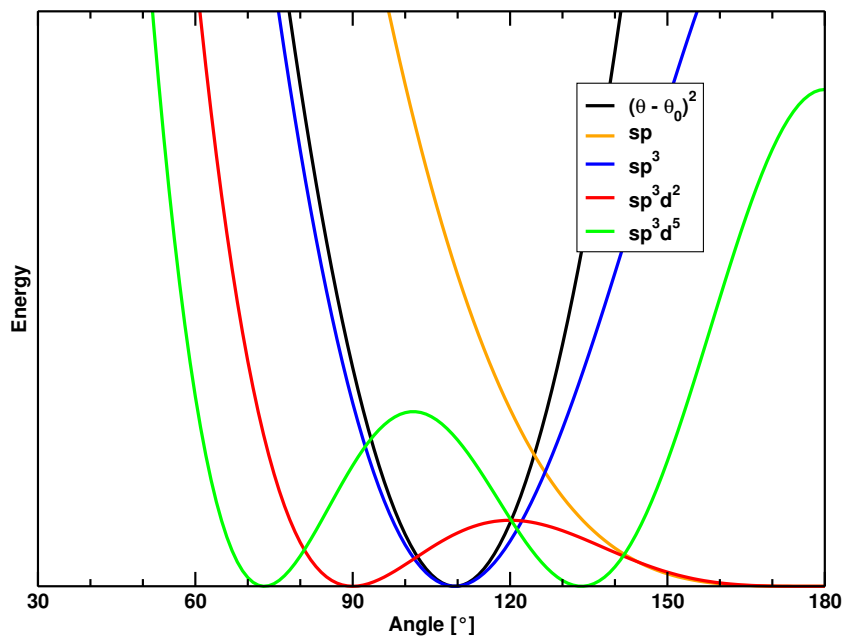$$E_{hyp} = k(1 - \Delta(\alpha + \pi)^2) \tag{2.27}$$

Figure 2.4.: The Valbond energy (arbitrary scale) as a function of the angle for several hybridizations (colours). For comparison, the energy function of a harmonic angle with $\theta_0 = 109.5°$ is shown as well (black).

The angle energy is the sum of all the nonhypervalent and hypervalent bonding orbital energies weighted by a bond order factor (BOF):

$$E_{angle} = \sum \text{BOF}_\text{i} \times E_{hyp} + \sum \text{BOF}_\text{j} \times E_{nonhyp} \tag{2.28}$$

The BOF is calculated from the assignment of the 3c4e bonds. Normal (nonhypervalent) orbitals have the weight 1 whereas hypervalent orbitals have a weight of 0.5 as the 3c4e bond is distributed over two normal bonds. The weight is distributed on all angles involving the specific orbital: e.g., if an orbital is involved in two angles, the weight is divided by these two orbitals to yield the BOF.

In hypervalent molecules there are several resonance structures composed of all possible assignments of the 3c4e bonds. The energy of each of the resonance structures is calculated and weighted. The weight of a resonance structure, described as the weighting

factor c_i, depends on the geometry and is expressed as:

$$c_i = \frac{\prod\limits_{i=1}^{hype} \Delta_i^2}{\sum\limits_{j=1}^{res} \prod\limits_{i=1}^{hype} \Delta_i^2} \tag{2.29}$$

The product of the overlapped orbitals for a certain resonance structure is weighted by the sum of the product for all resonance structures.

Finally, the total energy of this angle is depending on each of the n resonance structures and their weight c_i and energy E_i:

$$E_{\text{tot}} = \sum_i^n c_i E_i \tag{2.30}$$

One major advantage of Valbond over conventional force fields is that it requires considerably fewer parameters: only one parameter is required for every pair of bonded atoms. The remaining parameters (bonds, dihedral, nonbonding, etc.) are identical to those used in the standard CHARMM force field. However, partial optimizations of these parameters is likely to be necessary for a robust force field including specific metals in their different oxidation states. In addition, assignment of a hybridization is not always trivial and needs careful exploration for more complicated cases such as the piano stool complexes in the present study. In our implementation of VB and VBT[95] (available from CHARMM[71] version 37 onward), the conventional CHARMM angle terms are replaced by more realistic combinations of hybrid orbitals which can be assigned by the user. In this way, VB and VBT can be combined and are compatible with the conventional CHARMM force field and simulations in a variety of ways are possible.

2.6. Docking

Protein-ligand docking aims to predict and compare the structure of molecules interacting with a protein of known 3D structure.[134] Protein-ligand docking is a highly active field because of its applications in life sciences. Manual docking thanks to human intuition can yield very good structures as demonstrated in the even more difficult problem of protein folding with Foldit,[135,136] but when concerned with thousands of structures, only automated methods are viable. Popular docking programs include AutoDock[137] and AutoDock Vina[138], Glide[139], GOLD[140] or FlexX[141].

Most docking programs rely on the lock-key and induced-fit concepts. Different poses of the receptor are tested and a score resulting from the interaction of the molecules is calculated. These scoring functions are used to rate the validity of a docked structure. Typically, the output does not consist of a single docked structure, but of an ensemble of structures with a high score. The fact that proteins are in constant motion between different conformational states with similar energies is still often disregarded by docking programs, although some programs like AutoDock can employ flexible residues i.e. sidechains.

Different scoring functions are employed to yield the docked structures. One group relies on force fields to calculate the score from the intermolecular energy between the ligand and the receptor and the intramolecular energy of the ligand. GOLD and AutoDock are examples featuring a force field based scoring function. Knowledge-based scoring functions focus on statistical analysis of experimentally determined structures instead of trying to calculate binding effects explicitly. These potentials are based statistics of atom-atom pair and other interactions in large datasets of protein-ligand complexes of known structure. Some programs like FlexX combine several methods to yield a score.

For metals and metal complexes there is only very basic support. If available, it is usually limited to the most common metals in proteins like iron and zinc.[142] But support for other metals is practically absent.

The aim of most docking programs is to screen a lot of structures in limited time. But to do this, a lot of approximations are applied, false positive or false negatives do emerge and subsequent analysis with more exact methods is required. Molecular dynamics simulation is one of them.

2.7. Molecular Dynamics

Molecular dynamics (MD) simulations play an important role in biochemical research and material sciences. It can be used to model the motion of molecular systems like proteins at atomistic level. Molecules and atoms interact with each other and changes are propagated through the system as time advances. Trajectories for the atoms are determined by solving Newton's equations of motion for the system, forces acting on the particles are calculated by QM or MM, usually a force field is applied for systems with a large number of atoms. A general MD algorithm proceeds by iteratively calculating the acceleration a_i acting on an atom i from its mass m and forces F_i using Newtons second law of motion

$$a_i(t) = \frac{F_i\left(x(t)\right)}{m_i} \tag{2.31}$$

New coordinates at time $t + \Delta t$ are calculated with

$$x_i(t + \Delta t) = x_i(t) + v_i(t)\Delta t + \frac{a_i(t)\Delta t^2}{2} + O(\Delta t^3) + \cdots \tag{2.32}$$

Because a molecular systems consist of a vast number of particles and interactions, the forces and interaction energies are calculated using numerical integration in a (time)step-by-(time)step wise manner. It is necessary to account for time-reversibility and commonly used is a Verlet type[143] integrator. In a molecular dynamics system the total energy is always partitioned in a kinetic part from the movement of the atoms, and a potential part from the energy stored in the structure. The time-step for the integration is limited

by the fastest motion in the system to remain stable. Bonds involving hydrogens are the fastest motions in a typical MD simulation, but very often one is not interested in the exact simulation these fast vibrations but rather on slower processes. Time steps can be increased by using algorithms like SHAKE[144] to fix these fast vibrations.

For large systems involving slow motions, e.g. proteins, the longer a trajectory, the more meaningful results can be obtained as statistical noise for averages is reduced and also slow transitions can be sampled. The advance of computational power has opened up the possibility of doing more and more detailed calculations on larger and larger time scales. Massive parallelization of the simulation enables calculations in the nano to microsecond scale and using special purpose built computers like Anton, even simulations of several microseconds are accessible.[145,146]

Simulations can be performed in various ensembles. In the microcanonical or NVE ensemble, the number of atoms (N), the volume (V) and the total energy (E) are kept constant, whereas temperature and pressure are subject to change. It corresponds to a natural ensemble for MD as energy is conserved and no other constraints are in place.

In the canonical or NVT ensemble, the number of atoms (N), the volume (V) and the temperature (T) are kept constant, whereas total energy and pressure are subject to change. To yield a NVT ensemble, a thermostat is required to keep the temperature stable. The NVT ensemble is often used for the equilibration of a system.

The isothermal–isobaric or NPT ensemble is closely related to "real" experimental conditions in a open flask, as the number of atoms (N), the pressure (P) and the temperature (T) are kept constant, whereas total energy and volume are subject to change. A thermostat and a barostat is needed to keep these properties constant.

Solvent Considerations Many systems of interest in chemistry and biology and also the systems examined in this thesis are solvated. Whereas many metal complexes are soluble in organic solvents like methanol, biological systems are typically solvated in water.

Water modelling is thus an important characteristic in the description of the system. There are two general ways to consider solvent effects, either the solvent molecules are simulated explicitly, or the solvent is considered implicitly using a solvent model like the Poisson Boltzmann or Generalized Born model.[147] Using an implicit solvent reduces the cost for calculation drastically, but effects like the viscosity or the hydrophobic effect are not accounted for. Thus, explicit solvation is typically preferred. Several water models for solvents have been developed, the most widely used and the standard model for CHARMM is TIP3P.[148] To simulate a bulk system, simulation is usually performed using periodic boundary conditions, that is the system is replicated in such a way that particles that exit on one face of the system enter on the opposite face with the same velocity. Forces and interactions are replicated in the same way and enough solvent is needed to minimize effects of the solute on it self. As a result, a fully solvated system consists of many more atoms from the solvent than atoms from the solute.

MD simulations are a well established approach for the development of enzyme inhibitors and the general understanding of protein ligand interactions. In a similar way, we want to employ MD simulations to model metal-mediated protein-ligand interactions as required for the design and understanding of artificial metalloenzymes. However, although metalloproteins account for nearly half of all proteins in nature, computational modelling of protein-metal interactions is understudied and molecular mechanics programs and force field parameters compatible to proteins and transition metal are not readily available. Within this thesis, various computational approaches were pursued to model metal-mediated protein-ligand interactions. Several challenges were identified and thus addressed in the following chapters.

Part II.

Applications

3. Properties of Enzymes

The Protein Data Bank (PDB[149]) is a great library that facilitates studies on proteins and enzymes. Around two years ago, when the we performed our study on facial triads[150], 70'000 structures had been collected. Until now (beginning of 2013), it contains structural information of almost than 90'000 proteins, with most of the structures, 75'000, solved by X-ray. Most of the remaining structures were solved by NMR, only a few were determined using other means.[a] As the database continues to grow, the information contained within is multiplied as well as. The vast knowledge compressed within is useful for detailed analysis as well as for broad approaches to assess similarities and differences between proteins and thus the understanding of "how nature works". The information contained in the PDB has been widely used for all kind of analysis and numerous applications: From broad studies like counting zinc, iron or copper proteins in the human genome,[19,151–153] 3D motif search,[154] testing of scoring functions,[155] classification methods,[156,157] to narrow applications such as using a structure as starting point for MD simulations as done in this thesis, the PDB is a very valuable tool.

We set out to scrutinize the protein data base (PDB) to identify promiscuous metal binding sites. The following paper presents the work on the identification of latent facial-triad motifs amenable for the introduction of a transition metal. Upon insertion of the metal, the identified structures would become an artificial metalloenzyme. Metal-containing proteins available from the protein data bank were analyzed and prototypical facial-triad motif templates identified. After screening for these templates, metal binding

---

[a]http://www.rcsb.org

sites featuring a potential facial triad were proposed. The findings have been published in Metallomics.[150]

3.1. Identification of two-histidines one-carboxylate binding motifs in proteins amenable to facial coordination to metals

two-histidines one
carboxylate motif
"facial triad"

protein bearing a
latent facial triad
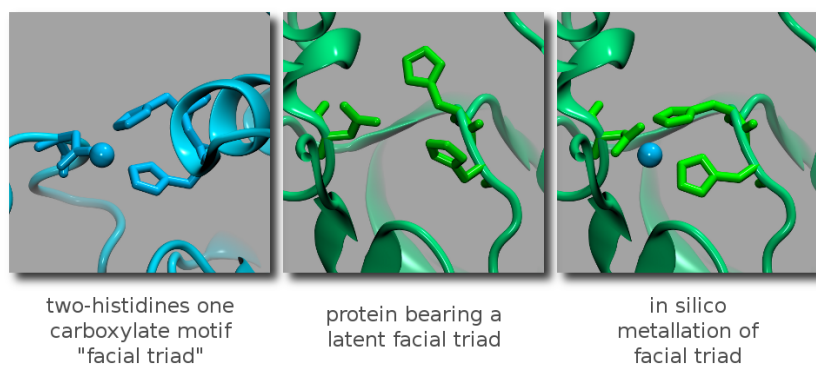
in silico
metallation of
facial triad

Figure 3.1.: Proteins bearing a latent facial triade were identified from prototypical two-histidines one-carboxylate binding motif templates. TOC figure of the article.[150]

Metallomics

PAPER

Identification of two-histidines one-carboxylate binding motifs in proteins amenable to facial coordination to metals†

Beat Amrein,‡[a] Maurus Schmid,‡[a] Guillaume Collet,[b] Philippe Cuniasse,[b] François Gilardoni,[c] Florian P. Seebeck*[d] and Thomas R. Ward*[a]

Among natural metalloenzymes, the facial two-histidines one-carboxylate binding motif (FTM) is a widely represented first coordination sphere motif present in the active site of a variety of metalloenzymes. A PDB search revealed a total of 1685 structures bearing such FTMs bound to a metal. Sixty statistically representative FTMs were selected and used as template for the identification of structurally characterized proteins bearing these three amino acids in a propitious environment for binding to a transition metal. This geometrical superposition search, carried out using the STAMPS software, returned 2320 hits. While most consisted of either apo-FTMs or bore strong sequence homology to known FTMs, seven such structures lying within a cavity were identified as novel and viable scaffolds for the creation of artificial metalloenzymes bearing an FTM.

Introduction

Metals are present in nearly half of the characterized proteome.[1] To carry out some of its most challenging transformations, Nature often relies on transition metals as cofactors.[2] Thanks to Darwinian evolution, the performance of many metalloenzymes approaches perfection both in terms of activity and selectivity.[3,4] Among the potential metal binding amino acid side chains, histidine and carboxylate (*i.e.* glutamate/aspartate) are the most prevalent.[5] For enzymes relying on mononuclear metal cofactors, the so-called *Facial Triad Motif*—combining two histidines and one carboxylate (either aspartate or glutamate, *fac*-[M(His$_2$)(O$_2$CR)], FTM hereafter)—has emerged as a recurring and versatile motif, Fig. 1.[6–8]

The broad functional scope of protein-bound metals found in naturally evolved catalysts has inspired a number of different approaches to create artificial metalloenzymes in the past twenty years.[9–18] Some of the most promising approaches in enzyme design combine *in silico* modelling with directed evolution protocols.[19–21] Creating catalytic function from a non-catalytic

scaffold remains challenging however.[22,23] To alleviate this, introduction of a metal cofactor offers an attractive means to generate novel functionality within a protein scaffold.[11,24] In this context, several successful strategies have been pursued: (i) covalent conjugation of a catalytically competent organometallic moiety;[25,26] (ii) supramolecular anchoring of a metal complex by exploiting a strong protein inhibitor interaction[27] and (iii) dative anchoring of a metal to amino acid side chains,[11,28] either present in the native protein or introduced by site-directed mutagenesis. In a "catalytic promiscuity" spirit,[29–31] we speculated that non-metal containing proteins may harbor potential metal binding sites which, upon addition of a suitable metal, may dramatically alter the function of the protein. Considering the wide occurrence of the *fac*-[M(His$_2$)(O$_2$CR)] motif, we set out to scrutinize the protein data base (PDB) to identify pre-organized two-histidines one-carboxylate triads in proteins for the facial coordination of transition metals.

Results and discussion

The identification of latent facial triad motifs was carried out in three steps: (1) identification of all structurally characterized *fac*-[M(His)$_2$(O$_2$CR)] motifs in the PDB; (2) selection of representative FTMs for metal coordination from the above dataset and (3) identification and classification of latent FTM motifs from metal-free proteins using the STAMPS software. Computational details are described in the methods section and in the ESI.†

Identification of *fac*-[M(His)$_2$(O$_2$CR)] motifs in the PDB

In the first step of the analysis, we screened the PDB (as of Oct 27, 2010) in search of metal-coordinated facial two-histidines one-carboxylate motifs (*fac*-[M(His)$_2$(O$_2$CR)]).

[a] *Department of Chemistry, University of Basel, Spitalstrasse 51, CH 4056-Basel, Switzerland. E-mail: thomas.ward@unibas.ch*
[b] *Service d'Ingénierie Moléculaire des Protéines, Institut de Biologie et Technologies de Saclay (IBITEC-S), Commissariat à l'Énergie Atomique (CEA), 91191 Gif-sur-Yvette, France*
[c] *FONGIT/FONGIT SEED INVEST, 18 Ch. des Aulx, CH 1228-Plan-les-Ouates, Switzerland*
[d] *Department of Chemistry, University of Basel, St-Johann Ring 19, CH 4056-Basel, Switzerland. E-mail: florian.seebeck@unibas.ch*
† Electronic supplementary information (ESI) available: Open-Structure scripts, the list of identified metalloenzymes (in PDB) containing a facial triad motif and the list of FTMs identified using the STAMPS software. See DOI: 10.1039/c2mt20010d
‡ Both authors contributed equally.

3. Properties of Enzymes
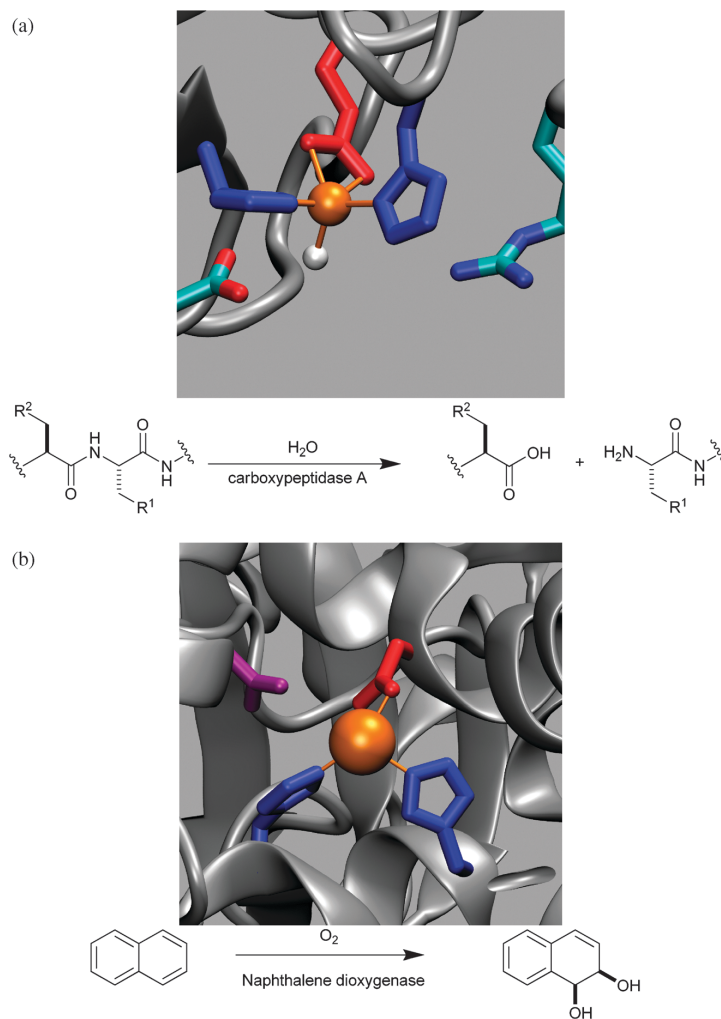
Fig. 1 Selected examples of metalloenzymes relying on a facial two-histidines one-carboxylate triad: carboxypeptidase A (a) and naphthalene dioxygenase (b).

From the initial 68 562 entries compiled from the PDB, 14 332 contained an explicit transition metal. In order to identify the FTM, a stepwise analysis was carried out using OpenStructure[32] scripts (see ESI†). The refinement strategy is presented in Scheme 1.

The dataset with the 14 332 structures containing explicit metal ions was further refined to exclude binuclear cofactors by setting a 4.5 Å threshold around the first transition metal. Next, only those entries containing two histidines and at least one carboxylate (either Glu or Asp) within 3.01 Å of the transition metal were selected, affording a total of 1429 hits. To exclude meridional triads mer-[M(His)$_2$(O$_2$CR)], the sum of the three angles spanned by the triad was determined (w_{tot}). Only those with $w_{tot} < 303°$ were selected, thus ensuring facial coordination for octahedral, square pyramidal and trigonal bipyramidal geometries, Fig. 2.[33] This filtering procedure

yielded a total of 1685 triads contained in 805 entries (*i.e.* more than one FTM per entry). The PDB entries containing >2 histidines bound to the transition metal were excluded. Those containing 2 histidines and ≥ 2 carboxylates were treated as independent hits, Fig. 1c.

The resulting FTM structures were classified according to the transition metal and binding motif, Fig. 3a and b respectively.

The presence of two histidines and one carboxylate gives rise to six different binding possibilities, depending on which N_{His} is bound and the nature of the carboxylate (Asp or Glu), Fig. 3b. Analysis of the binding motifs revealed a large majority of fac-(Nε, Nε, Asp) motifs (1006 hits), followed by fac-(Nε, Nε, Glu) (485 hits). It is interesting to note that both the (Nδ, Nδ, Glu) and (Nδ, Nδ, Asp) are widely represented in the mer-[M(His)$_2$(O$_2$CR)] structures but nearly absent in the

46
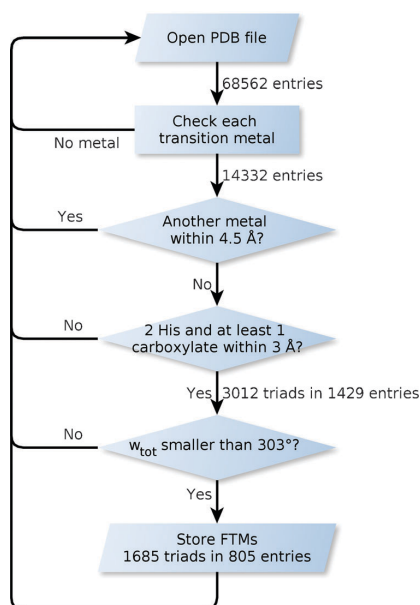
Scheme 1 Flowchart documenting the identification of structurally characterized *fac*-[M(His)$_2$(O$_2$CR)]. w_{tot} corresponds to the sum of the L–M–L angles spanned by the facial triad motif, see Fig. 2.

fac-[M(His)$_2$(O$_2$CR)] geometry, Table 1. Concerning the transition metal bound to the FTM, zinc (despite exclusion

of the tetrahedral geometry based on the w_{tot} criterion), iron and manganese were the most abundant metals with 967, 315 and 165 representatives respectively, Fig. 3a. A list of all PDB structures containing an FTM is collected in the ESI.†

Several noteworthy features arise from this analysis: (i) for a large majority of the identified *fac*-[M(His)$_2$(O$_2$CR)] motifs, the three coordinating amino acids are situated on a single polypeptide chain (96.4% of all entries) and (ii) a monodentate binding mode of the Asp/Glu side chain dominates the structurally characterized *fac*-[M(His)$_2$(η^1-O$_2$CR)]. Setting a $\delta \leq 0.15$ Å threshold between the two M–O distances yields only 50 hits (3.0% of the 1685 FTM motifs identified). This suggests that, although catalytically relevant, the *fac*-[M(His)$_2$(η^2-O$_2$CR)] coordination mode is rarely achieved under the conditions required for crystallization.

Selection of representative FTMs for metal coordination

With the aim of identifying the most representative FTMs among the 1685 structures returned from the PDB search, a set of 24 descriptors for each structure was subjected to Principal Component Analysis (PCA). These descriptors correspond to the 24 linearly independent internal coordinates (*i.e.* $3N - 6$) spanned by the 10 atoms which make up the FTM for each structure: M, N$_{His}$, Cα_{His}, Cβ_{His}, N$_{His'}$, C$\alpha_{His'}$, C$\beta_{His'}$, O$_{Glu/Asp}$, C$\alpha_{Glu/Asp}$, C$\beta_{Glu/Asp}$, Fig. 2d. The PCA allowed reduction of the descriptor space to nine dimensions ($Q^2 = 0.974$, cross validated). Cross-terms were added to enhance the performance of the PCA. A 3D projection of all FTM hits in PCA space (PCA1, PCA2 and PCA3) is depicted in Fig. 4.
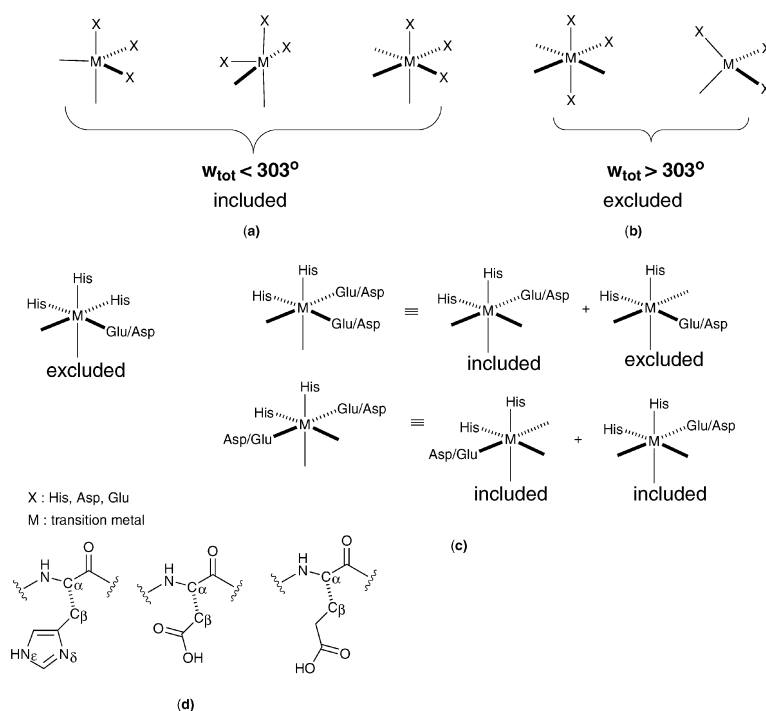
Fig. 2 Possible geometries spanned by *fac*-[M(His)$_2$(O$_2$CR)] (a–c); and atom definition used in this study (d).

47
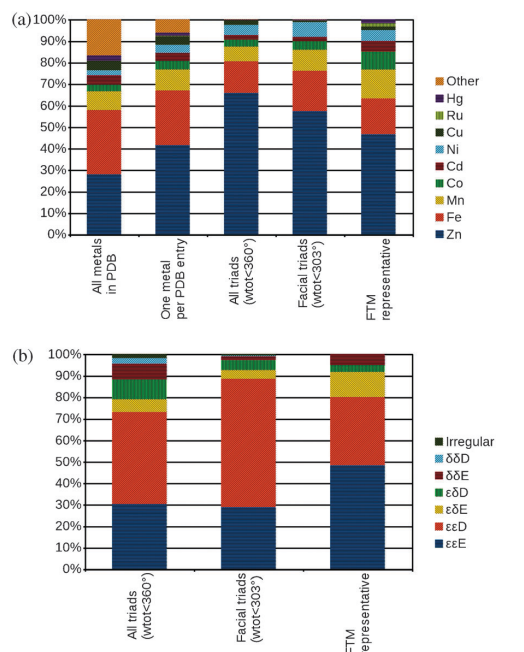
Fig. 3 Classification of the *fac*-[M(His)$_2$(O$_2$CR)] according to metal (a); and coordinating side chains (b).

Fig. 4 Projection in the PCA1, PCA2 and PCA3 space of all FTMs identified in the PDB search (black spheres) and FTMs selected by *D*-optimal design and used as template for the STAMPS search (red spheres).

Table 1 Summary of coordination motifs for [M(His)$_2$(O$_2$CR) identified in the PDB search

Entry		εεE	εεD	εδE	εδD	δδE	δδD	irr.[a]	Total
1	All triads	911	1288	179	277	222	79	56	3012
2	FTM $w_{tot} \leq 303°$	485	1006	69	78	29	6	12	1685
3	STAMPS search	29	19	7	2	3	0	0	60

[a] irr: irregular motifs which could not be unambiguously assigned.

Next, the *D*-optimal design procedure was used to select the most representative subset of FTMs from the initial library of 1685 structures.[34–36] This statistical analysis allowed the identification of a subset of 60 structures which carried the most diversity in the PCA space. The hits are listed in Table 2 and displayed in the PCA space in Fig. 4. These structures were used as template for the STAMPS software search.

Identification and classification of latent FTM motifs from metal-free proteins

Several algorithms for the prediction of metal binding sites have been reported. Some of these rely on: (i) primary sequence alignment;[37,38] (ii) structure-based methods,[39–43] (iii) force-fields,[44] or (iv) combinations thereof.[45,46] With the aim of identifying *novel protein scaffolds* not related to any known FTM scaffolds, we scrutinized the local environment spanned by the three amino acids forming the first coordination sphere around the metal (*i.e.* His, His and Glu or Asp). For this purpose, we relied on the Search for Three dimensional Atom Motifs in Protein Structures (STAMPS) software.[47,48] This program is also available through the web-site RASMOT-3D

PRO and offers a convenient interface to search the PDB for a user defined motif and to detect related patterns (based on the relative positions of the Cα$_{His}$, Cβ$_{His}$, Cα$_{His'}$, Cβ$_{His'}$, Cα$_{Glu/Asp}$ and Cβ$_{Glu/Asp}$) in the database. However, due to some limitations in the RASMOT-3D PRO (maximum number of returned structures limited to 100 and lower flexibility in the selection of equivalent residues), the motif search was carried out with the developmental version of STAMPS.

Each of the 60 FTM motifs identified by the *D*-optimal design listed in Table 2 was used as template for the STAMPS search. The STAMPS search allowed the identification of characterized structures bearing either a His-His-Glu or a His-His-Asp motif in a similar geometric arrangement to that of one of the 60 template motifs.

With the aim of narrowing down the potential FTM candidates, a selection procedure, summarized in Scheme 2, was applied to the identified structures. From the 9012 solutions returned by STAMPS, 4018 were present in the hits identified from Scheme 1. Additionally, a transition metal within 7.5 Å (for Glu), 6.2 Å (for Asp), 9.3 Å (for His) of a Cα was found in 2674 cases. These distances are the maximum distances between the metal and the Cα of the respective residues observed in the crystal structure. This procedure afforded 2320 potential FTM motifs (located on 1368 chains).

Having removed natural FTMs and other metal containing proteins, the fasta files from the remaining 1368 PDBs were filtered through blastclust (http://toolkit.tuebingen.mpg.de) to eliminate redundancy that emerges from the same protein being crystallized in many different formats. This procedure reduced the list to 418 PDB files (18%).

Next, each of the PDB files was inspected visually to identify and evaluate the first and second coordination sphere around the His and Glu/Asp residues. The resulting potential FTMs were classified as either (i) putative, (ii) poor or (iii) viable binding sites (see ESI† for a complete list containing all PDB files and their classification).

(i) Putative facial triad motifs

Nearly one third (127 entries) of the proteins returned by the STAMPS software were identified as either apoproteins of

48

Table 2 Most representative facial triad motifs identified by PCA and D-optimal design

Entry	PDB code	Chain	Metal	Binding mode	#hits[a]
01	1ARM	A	Hg	δδE	202
02	1B4U	D	Fe	εεE	38
03	1BK0	A	Fe	εεD	314
04	1C0W	B	Co	εδE	83
05	1C0W	D	Co	εδE	73
06	1C0W	D	Co	εδE	94
07	1CJX	C	Fe	εεE	207
08	1DHY	A	Fe	εεE	173
09	1DRY	A	Fe	εεE	184
10	1E5H	A	Fe	εεD	314
11	1EIQ	A	Fe	εεE	184
12	1F5T	D	Ni	εδE	275
13	1G8G	B	Cd	εεD	115
14	1G8H	A	Cd	εεD	109
15	1GW6	A	Zn	εεE	552
16	1HA5	D	Zn	εδE	53
17	1IM5	A	Zn	εεD	30
18	1KHO	A	Zn	εεE	69
19	1OE2	A	Cu	εεE	452
20	1OI0	A	Zn	εεD	29
21	1TU4	B	Co	δδE	9
22	1U8R	B	Co	εδE	99
23	1XN0	A	Zn	εεD	281
24	1Y8J	A	Zn	εεE	235
25	1ZZ7	A	Fe	εεE	223
26	1ZZ8	B	Fe	εεE	111
27	2CQZ	B	Ni	εεD	393
28	2EIK	C	Cd	εεE	30
29	2EIN	L	Zn	εεE	123
30	2F1D	E	Mn	εεE	46
31	2FGE	B	Zn	εεE	162
32	2FLI	K	Zn	εδD	253
33	2FM0	A	Zn	εεD	288
34	2HXG	B	Mn	εεE	39
35	2HXG	C	Mn	εεE	28
36	2IT0	B	Ni	εδE	89
37	2O3Z	A	Zn	εεD	58
38	2O3Z	B	Zn	εεD	52
39	2QYN	B	Zn	εεE	302
40	2V8J	A	Mn	εεE	27
41	2WBY	A	Zn	εεE	192
42	2WC0	B	Zn	εεE	206
43	2ZXC	B	Zn	εεE	8
44	3B4R	B	Zn	εεD	78
45	3BG3	C	Mn	εεD	91
46	3BG5	B	Mn	εεD	100
47	3BKL	A	Zn	εεE	48
48	3D19	E	Fe	εεE	66
49	3E4A	B	Zn	εεE	185
50	3E4Z	B	Zn	εεE	184
51	3EF7	B	Zn	εεE	40
52	3FXS	A	Ru	εεE	1379
53	3HK7	J	Zn	εεD	294
54	3HK9	J	Zn	εεD	383
55	3HM7	C	Zn	εεD	266
56	3HO8	B	Mn	εεD	105
57	3HWP	A	Zn	εεE	45
58	3Mn8	A	Zn	δδE	207
59	3N9D	A	Mn	εδD	106
60	3OME	B	Zn	εεE	148

[a] This corresponds to the number of hits returned when using the considered entry as template for the STAMPS search.

known metalloenzymes (example: TauD apo structure PBD: 1OTJ compared to PBD: 1OS7, Fig. 5) or proteins with unknown functions but bearing a clear homology to known metalloproteins such as the apo form of a proline 3-hydroxylase (PBD 1E5R), the conserved protein YecM from *Escherichia coli*

(PBD 1K4N), or a member of the glyoxalase superfamily (PBD 3IUZ).

(ii) Poor facial triad motifs

Upon visual inspection, one third of the identified protein structures (163 entries) appeared as unlikely metal binding sites. The structures were classified as poor facial triad motifs if at least one of the following characteristics applied:

(a) the three triad residues are located such that none of their conformers could possibly generate a facial tridentate metal binding site (example: PBD 1VK5, Fig. 6a).

(b) at least one of the triad residues is engaged in a tight hydrogen-bonding network or a buried salt bridge. For example, a glutamate side chain, which forms a buried salt bridge with an arginine side chain, is an unlikely ligand for a metal (example: PDB 1HXA, Fig. 6b).

(c) the potential metal binding site is occupied either by other side chains (example PDB 2ICH, Fig. 6c) or the peptide backbone.

Among the structures with poor metal binding potential, we found many ankyrin repeat proteins (16 entries out of 163; for example: 2DVW, 1YYH, 1MJ0, 1TR4). The positions of the identified two-His one-carboxylate triad are conserved among the 16 sequences. The three side chains are engaged in multiple hydrogen bonds with the peptide backbone and presumably play an important role in stabilizing the ankyrin repeat fold making metal binding by these proteins through this motif unlikely.

Viable. Triads that emerged as latent metal binding sites and without apparent homology to documented metal binding sites were classified as viable. The majority of those triads (114 out of 121) are located on the protein surface (Fig. 7a).

Given that histidine is one of the least frequent residues in proteins (2%), accidental pairs of histidines flanked by a neighboring aspartic- or glutamic acid are improbable. Therefore, these constellations are likely to hold functional significance. For example, in the presence of high concentrations of earth-alkali (Mg^{2+}, Ca^{2+}) and even transition metals (Mn^{2+}) as found in certain cell types or in extracellular medium,[49] low affinity metal binding to the protein surface may tune the protein's stability, activity or aggregation state.

We also found seven surface binding motifs which include histidines from the N- or C-terminal His-tag of recombinant proteins underscoring that His-tags can play non-innocent roles in the creation of promiscuous metal binding sites.

The initial goal of our search was to identify undocumented metal-binding sites within structurally characterized proteins that might confer a moonlighting function as a metalloprotein and/or would provide virgin ground to design novel artificial metalloenzymes.[50,51]

From the initial set of 51 512 PDB files, we identified 114 surface borne FTMs (for example: PBD 1BM0, Fig. 7a) as well as seven FTMs located within a cavity (Fig. 7b–h). No significant commonalities are apparent among these seven structures, except that they fit the above FTM search scheme. Five of the proteins are enzymes, which probably reflects the over-representation of catalysts in the PDB. Further evaluation of their ability to bind metal ions and to unravel their catalytic potential will have to await *in vitro* characterization.

3. Properties of Enzymes
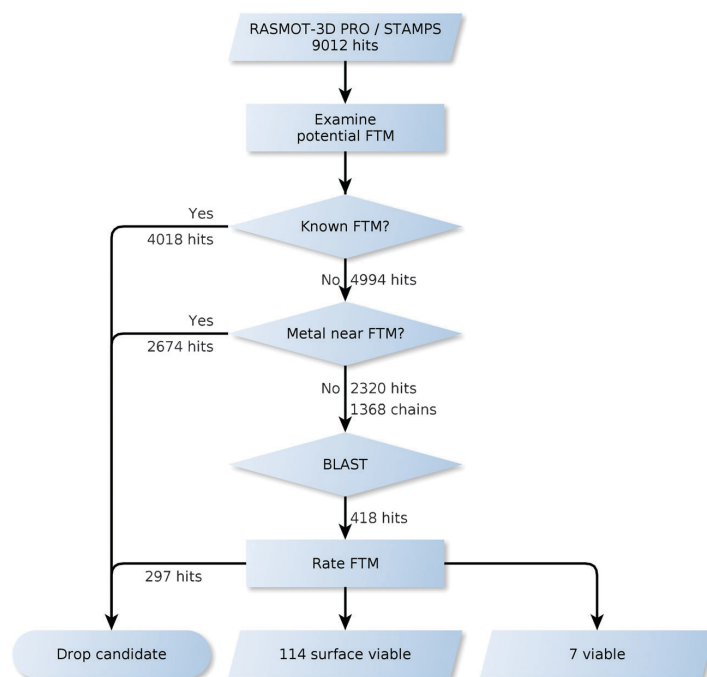
Scheme 2 Flowchart documenting the identification of potential two-histidines one-carboxylate motifs in the non-redundant chain set.
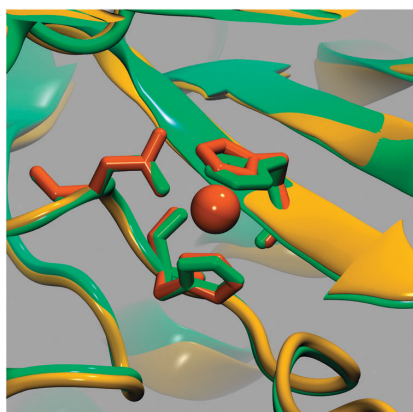
Fig. 5 Putative facial triad motifs identified by STAMPS: a TauD apoprotein PBD 1OTJ (green) and its metallated congener PBD 1OS7 (orange).

This observation raises the question as to whether these seven potential FTMs buried within a "binding pocket" have a cryptic biological function. If functionally significant, we would expect these motifs to be conserved among closely related protein sequences. However, in four of the seven sequences at least one of the triad residues is not conserved among homologs with at least 60% sequence similarity, suggesting that those potential metal binding motifs are probably not functionally relevant (Table 3). Unnecessary metal affinity in protein cavities such as enzyme active sites could lead to metal inhibition. Therefore, we would expect that such a trait is subject to a negative evolutionary pressure. This is consistent with our finding that undocumented metal binding sites are far more common on protein surfaces (114) than in cavities (7). On the other hand, few mutations in the primary sequence of these seven proteins might give rise to novel metal binding sites and potential catalytic function. These thus present fertile ground for the emergence of metal-dependent catalytic activity.

Methods

Identification of *fac*-[M(His)$_2$(O$_2$CR)] motifs in the PDB

In the first step of the analysis, we screened the PDB (as on Oct 27, 2010) in search of metal-coordinated facial two-histidines one-carboxylate motifs (*fac*-[M(His)$_2$(O$_2$CR)]). For this purpose, the analysis was carried out sequentially as summarized in Scheme 1.

For the first step of the analysis, the database was down-loaded to a local hard disk and every PDB entry was scanned for the presence of a transition metal and its first coordination sphere was analyzed as outlined in Scheme 1. In order to automate this geometrical analysis, OpenStructure scripts were implemented.[32] All scripts used for this automated search are collected in the ESI.†

Principal component analysis

A principal component analysis was carried out on the 24 linearly independent internal coordinates (*i.e.* $3N - 6$) spanned by the 10 atoms which make up the FTM of the identified 1685 structures: M, N$_{His}$, C(α)$_{His}$, C(β)$_{His}$, N$_{His'}$, C(α)$_{His'}$, C(β)$_{His'}$,

50
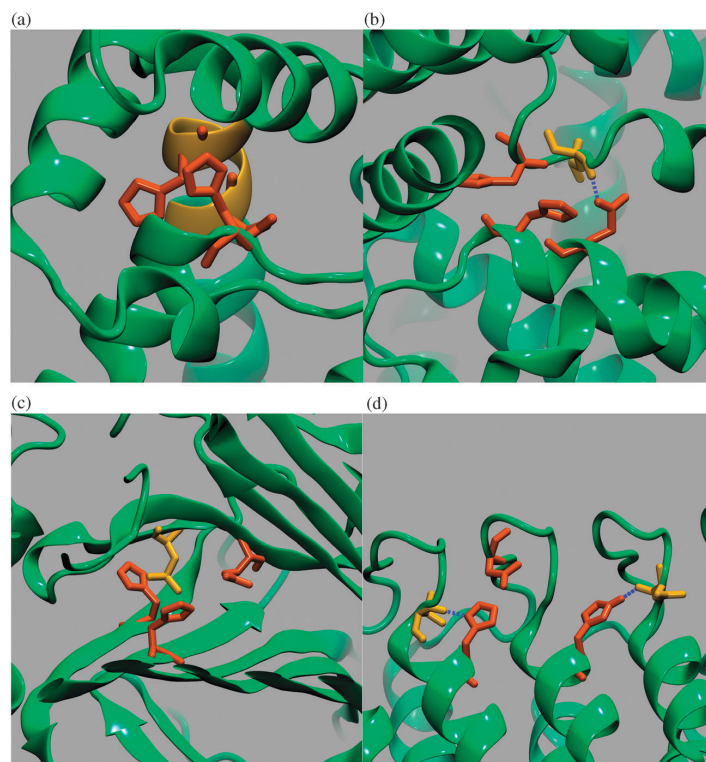
Fig. 6 Potential facial triad motifs identified by STAMPS classified as poor due to: steric clash of the triad (orange) with the protein's main chain (green, problematic region highlighted in yellow), PDB: 1VK5 (a); H-bonding (blue) or salt-bridges hampering creation of a propitious FTM, PDB: 1HXA (b) and steric clash of the triad with an amino acid side chain (yellow) with restricted mobility, PDB: 2ICH (c); ankyrin highlighting the multiple H-bonding network of the potential FTM, PDB: 2DVW.

$O_{Glu/Asp}$, $C(\alpha)_{Glu/Asp}$, $C(\beta)_{Glu/Asp}$, Fig. 2d. Based on this procedure, the original 24-dimensions descriptor space was reduced to nine dimensions ($Q^2 = 0.974$, cross validated). Cross-terms were added to enhance the performance of the PCA.

D-Optimal design

A computer aided design was used to select the most representative subset of FTMs from the initial library of 1685 structures. The D-optimal design was applied to sample the PC space and to identify proteins that carried the maximum diversity in the FTM library.[35] D-optimal designs are generated by an iterative search algorithm and seek to minimize the covariance of the parameter estimates—i.e. the Principal Components—for a quadratic model in the present case.

Both PCA and D-optimal design were computed using the R-project software package (freely available at http://www.r-project.org/).[35]

STAMPS search[47,48]

Given a triplet motif, STAMPS searches for residue triplets (i.e. the FTM) with a similar topology in the PDB (in our case, the non-identical set http://www.ncbi.nlm.nih.gov/Structure/VAST/nrpdb.html).[57] Comparison between the input motif

and the protein under scrutiny is carried out on $C\alpha$ and $C\beta$ atoms in two steps.

First, in a distance comparison routine, all interatomic distances between the $C\alpha$ and $C\beta$ atoms of the input motif are calculated. These distances, in our case the $C\alpha$ and $C\beta$ of both histidines and either the glutamate or the aspartate of the FTM, are compared with the interatomic distances between all $C\alpha$ and $C\beta$ atoms in the protein considered. A distance threshold (Δ) is specified for each of these distances. If this cutoff is exceeded for a given triplet, it is rejected. In this study, the cutoff was set to 1.0 Å.

The accepted triplets are subsequently tested with the second routine. A root mean square deviation (RMSD) filter is applied. After translation and rotation of the entire protein to minimize RMSD, triplets that have a RMSD of <1.0 Å for the $C\alpha$ and $C\beta$ atoms are retained.

This motif search in STAMPS was performed on the non-identical chain set containing 51 512 entries. As a consequence, only two-histidines one-carboxylate motifs located on the same polypeptide chain could be identified. In light of the results for naturally occurring facial triads (>96.4% on a single polypeptide chain), this limitation is acceptable. A version of the STAMPS software is available through the web-server RASMOT-3D PRO (http://biodev.cea.fr/rasmot3d/).

51

3. Properties of Enzymes
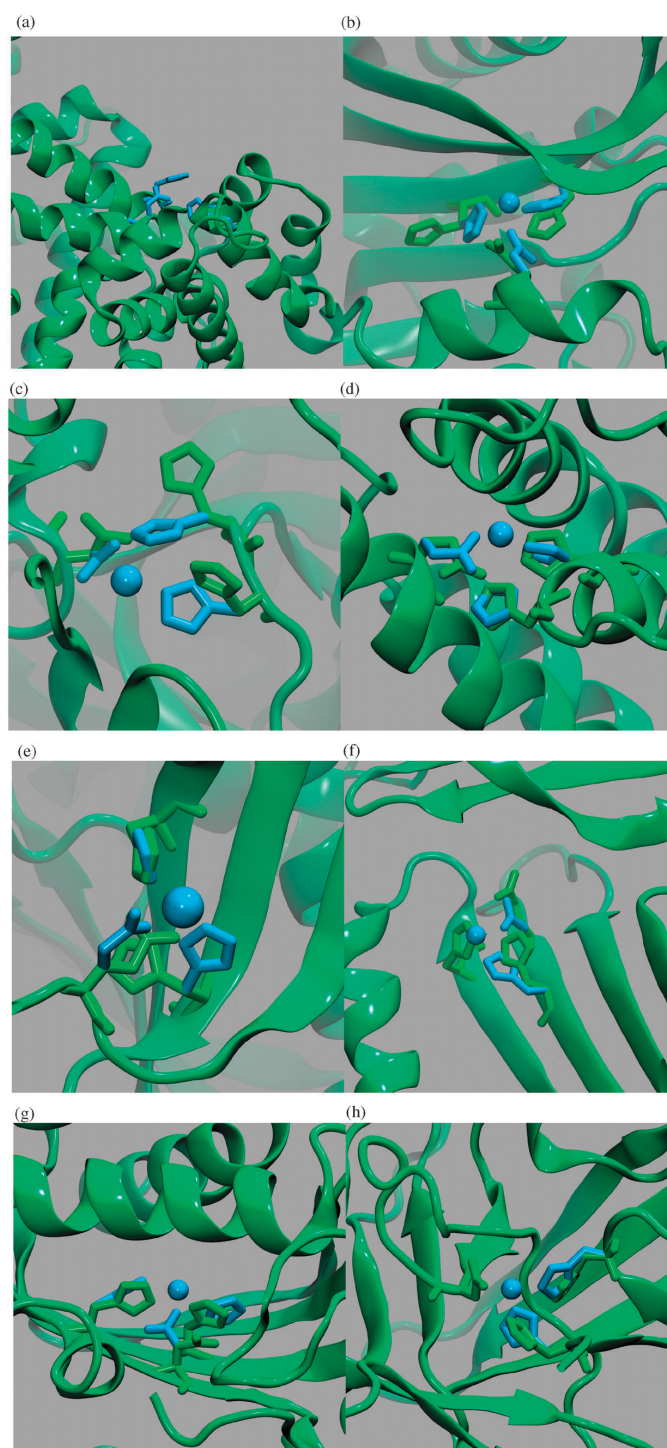
Fig. 7 Potential facial triad motifs identified by STAMPS classified as viable. Green: cartoon representation of the crystal structure and amino acid side chains spanning the triad. Blue: rotated amino acid side chains able to accommodate a hypothetical metal (blue sphere): located on the surface PBD 1BM0 (a), located in a pocket PBD 3A27 (b); PBD 3D53 (c); PBD 1GJV (d); PBD 3LOU (e); PBD 2W4Y (f); PBD 2F99 (g) and PBD 1JSY (h).

This journal is © The Royal Society of Chemistry 2012

52

Table 3 Summary of protein structures hosting latent metal binding sites as identified by the STAMPS software and classified by visual inspection. Residues which are not conserved among homologs with at least 60% sequence similarity are indicated by a *

PDB code	Annotated protein function	Carboxylate	Histidine 1	Histidine 2
3A27[52]	Aminocarboxypropyl-transferase	Glu86*	His193*	His195*
3D53	Pyrophosphatase	Asp114*	His76*	His77*
1GJV[53]	BCKD kinase	Glu118	His177	His181*
3LOU	Formyltetrahydrofolate deformylase	Glu236	His199	His228
2W4Y[54]	Caulobacter bacteriophage 5—virus-like particle	Glu60	His89	His62
2F99[55]	Polyketide cyclase AKNH	Asp121*	His107*	His119*
1JSY[56]	Bovine arrestin-2	Asp290	His210	His353

Outlook

The straightforward approach presented herein is a valuable tool for the identification of rare structural constellations. In the present case, the two-histidines one-carboxylate triad may, upon addition of a transition metal, give rise to radically new catalytic function to a given protein/enzyme (*i.e.* catalytic promiscuity). Importantly, the identified 121 potential facial triad motifs are based on first coordination sphere considerations only and not related to any previously reported FTM. We anticipate that this method may (i) prove valuable for mechanistic enzymology and (ii) allow the role of catalytic promiscuity in the evolutionary emergence of novel enzymatic activity to be probed. Current efforts in the group are directed at producing these seven proteins and evolving their catalytic properties upon addition of various transition metals.

Acknowledgements

References

1 Y. Lu, N. Yeung, N. Sieracki and N. M. Marshall, *Nature*, 2009, **460**, 855.
2 I. Bertini, A. Sigel and H. Sigel, *Handbook on Metalloproteins*, Marcel Dekker Inc, New York, 2001.
3 F. H. Arnold, *Acc. Chem. Res.*, 1998, **31**, 125.
4 W. P. C. Stemmer, *Nature*, 1994, **370**, 389.
5 C. Andreini, I. Bertini, G. Cavallaro, R. J. Najmanovich and J. M. Thornton, *J. Mol. Biol.*, 2009, **388**, 356.
6 P. C. A. Bruijnincx, G. v. Koten and R. J. M. K. Gebbink, *Chem. Soc. Rev.*, 2008, **37**, 2716.
7 K. D. Koehntop, J. P. Emerson and L. Que Jr., *JBIC, J. Biol. Inorg. Chem.*, 2005, **10**, 87.
8 G. Parkin, *Chem. Rev.*, 2004, **104**, 699.
9 T. R. Ward, *Acc. Chem. Res.*, 2011, **44**, 47.
10 T. Heinisch and T. R. Ward, *Curr. Opin. Chem. Biol.*, 2010, **14**, 184.
11 J. Podtetenieff, A. Taglieber, E. Bill, E. J. Reijerse and M. T. Reetz, *Angew. Chem., Int. Ed.*, 2010, **49**, 5151.
12 D. P. Weiner, T. Wiemann, M. M. Wolfe, P. Wentworth Jr and K. D. Janda, *J. Am. Chem. Soc.*, 1997, **119**, 4088.
13 B. I. Iverson, S. A. Iverson, V. A. Roberts, E. D. Getzoff, J. A. Tainer, S. J. Benkovic and R. A. Lerner, *Science*, 1990, **249**, 659.
14 *Catalytic Antibodies*, ed. E. Keinan, Wiley-VCH, Weinheim, 2005.
15 P. G. Schultz and R. A. Lerner, *Science*, 1995, **269**, 1835.
16 H. S. Lee and P. G. Schultz, *J. Am. Chem. Soc.*, 2008, **130**, 13194.
17 D. Coquière, J. Bos, J. Beld and G. Roelfes, *Angew. Chem., Int. Ed.*, 2009, **48**, 5159.
18 F. Rosati and G. Roelfes, *ChemCatChem*, 2010, **2**, 916.
19 L. Jiang, E. A. Althoff, F. R. Clemente, L. Doyle, D. Röthlisberger, A. Zanghellini, J. L. Gallaher, J. L. Betker, F. Tanaka, C. F. Barbas III, D. Hilvert, K. N. Houk, B. L. Stoddard and D. Baker, *Science*, 2008, **319**, 1387.
20 D. Röthlisberger, O. Khersonsky, A. M. Wollacott, L. Jiang, J. DeChancie, J. Betker, E. A. Althoff, A. Zanghellini, O. Dym, S. Albeck, K. N. Houk, D. S. Tawfik and D. Baker, *Nature*, 2008, **453**, 190.
21 J. B. Siegel, A. Zanghellini, H. M. Lovick, G. Kiss, A. R. Lambert, J. L. St Clair, J. L. Gallaher, D. Hilvert, M. H. Gelb, B. L. Stoddard, K. N. Houk, F. E. Michael and D. Baker, *Science*, 2010, **329**, 309.
22 B. Seelig and J. W. Szostak, *Nature*, 2007, **448**, 828.
23 H. S. Park, S. H. Nam, J. K. Lee, C. N. Yoon, B. Mannervik, S. J. Benkovic and H. S. Kim, *Science*, 2006, **311**, 535.
24 N. Yeung, Y.-W. Lin, Y.-G. Gao, X. Zhao, B. S. Russell, L. Lei, K. D. Miner, H. Robinson and Y. Lu, *Nature*, 2009, **462**, 1079.
25 D. Qi, C.-M. Tann, D. Haring and M. D. Distefano, *Chem. Rev.*, 2001, **101**, 3081.
26 P. J. Deuss, R. den Heeten, W. Laan and P. C. Kamer, *Chem.–Eur. J.*, 2011, **17**, 4680.
27 J. Steinreiber and T. R. Ward, *Coord. Chem. Rev.*, 2008, **252**, 751.
28 K. Okrasa and R. J. Kazlauskas, *Chem.–Eur. J.*, 2006, **12**, 1587.
29 M. D. Toscano, K. J. Woycechowsky and D. Hilvert, *Angew. Chem., Int. Ed.*, 2007, **46**, 3212.
30 O. Khersonsky, C. Roodveldt and D. S. Tawfik, *Curr. Opin. Chem. Biol.*, 2006, **10**, 498.
31 U. T. Bornscheuer and R. J. Kazlauskas, *Angew. Chem., Int. Ed.*, 2004, **43**, 6032.
32 M. Biasini, V. Mariani, J. Haas, S. Scheuber, A. D. Schenk, T. Schwede and A. Philippsen, *Bioinformatics*, 2010, **26**, 2626.
33 D.-H. Jo and J. Q. Jr., *Angew. Chem., Int. Ed.*, 2000, **39**, 4284.
34 M. Otto, *Chemometrics: Statistics and computer application in analytical chemistry*, Wiley VCH, Weinheim, 2nd ed, 2007.
35 F. Gilardoni, B. McKay, A. Graham and B. Brown, *Book of Abstracts of the 225th ACS National Meeting*, 2003.
36 Y. M. Yu, *Stat. Comput.*, 2011, **21**, 475.
37 A. Passerini, C. Andreini, S. Menchetti, A. Rosato and P. Frasconi, *BMC Bioinformatics*, 2007, **8**, 1471.
38 R. Thilakaraj, K. Raghunathan, S. Anishetty and G. Pennathur, *Bioinformatics*, 2007, **23**, 267.
39 J. S. Sodhi, K. Bryson, L. J. McGuffin, J. J. Ward, L. Wernisch and D. T. Jones, *J. Mol. Biol.*, 2004, **342**, 307.
40 D. S. Gregory, A. C. R. Martin, J. C. Cheetham and A. R. Rees, *Protein Eng., Des. Sel.*, 1993, **6**, 29.
41 R. Levy, M. Edelman and V. Sobolev, *Proteins: Struct., Funct., Bioinf.*, 2009, **76**, 365.
42 M. Babor, S. Gerzon, B. Raveh, V. Sobolev and M. Edelman, *Proteins: Struct., Funct., Bioinf.*, 2008, **70**, 208.
43 B. Seebeck, I. Reulecke, A. Kamper and M. Rarey, *Proteins: Struct., Funct., Bioinf.*, 2008, **71**, 1237.
44 J. W. H. Schymkowitz, F. Rousseau, I. C. Martins, J. Ferkinghoff-Borg, F. Stricher and L. Serrano, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 10147.
45 M. Brylinski and J. Skolnick, *Proteins: Struct., Funct., Bioinf.*, 2011, **79**, 735.
46 C. Andreini, I. Bertini, G. Cavallaro, G. L. Holliday and J. M. Thornton, *JBIC, J. Biol. Inorg. Chem.*, 2008, **13**, 1205.

3. Properties of Enzymes

47 G. Debret, A. Martel and P. Cuniasse, *Nucleic Acids Res.*, 2009, **37**, W459.

48 C. Magis, D. Gasparini, A. Lecoq, M. H. Le Du, E. Stura, J. B. Charbonnier, G. Mourier, J. C. Boulain, L. Pardo, A. Caruana, A. Joly, M. Lefranc, M. Masella, A. Menez and P. Cuniasse, *J. Am. Chem. Soc.*, 2006, **128**, 16190.

49 E. Demedicis, J. Paquette, J. J. Gauthier and D. Shapcott, *Appl. Environ. Microbiol.*, 1986, **52**, 567.

50 C. J. Jeffery, *Curr. Opin. Struct. Biol.*, 2004, **14**, 663.

51 J. Martinez-Fabregas, S. Rubio, A. Diaz-Quintana, I. Diaz-Moreno and M. A. De la Rosa, *FEBS J.*, 2011, **278**, 1401.

52 M. Umitsu, H. Nishimasu, A. Noma, T. Suzuki, R. Ishitani and O. Nureki, *Proc. Natl. Acad. Sci. U. S. A.*, 2009, **106**, 15616.

53 M. Machius, J. L. Chuang, R. M. Wynn, D. R. Tomchick and D. T. Chuang, *Proc. Natl. Acad. Sci. U. S. A.*, 2001, **98**, 11218.

54 P. Plevka, A. Kazaks, T. Voronkova, S. Kotelovica, A. Dishlers, L. Liljas and K. Tars, *J. Mol. Biol.*, 2009, **391**, 635.

55 P. Kallio, A. Sultana, J. Niemi, P. Mantsala and G. Schneider, *J. Mol. Biol.*, 2006, **357**, 210.

56 S. K. Milano, H. C. Pace, Y. M. Kim, C. Brenner and J. L. Benovic, *Biochemistry*, 2002, **41**, 3321.

57 J. F. Gibrat, T. Madej and S. H. Bryant, *Curr. Opin. Struct. Biol.*, 1996, **6**, 377.

54

4. Protein Ligand Simulations

The simulation of a protein-ligand system is first step towards the simulation of artificial metalloenzymes featuring transition metal complexes. Human carbonic anhydrase 2 as a potential scaffold for an artificial metalloenzyme was chosen as a model system and explored which is summarized in section 4.3. Docking studies on the established biotin-streptavidin system ilustrated the usefulness for computational investigations as demonstrated in chapter 5.

4.1. Carbonic Anhydrase as a Model Enzyme

Carbonic anhydrase (CA) is a well suited model system for studying enzymes. Its natural function is the catalyzation of the hydration of CO2 and the dehydration of bicarbonate.

$$CO_2 + H_2O \rightleftharpoons HCO_3^- + H^+ \tag{4.1}$$
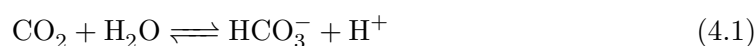
This well characterized enzyme has been the target of many studies. There are many isoforms of CA that have been discovered. The active site of CAs consists of a catalytic ZnII ion that is coordinated by three histidines and one water in a distorted tetrahedral geometry (see Figure 4.1). CA is particularly attractive for biophysical studies of protein-ligand binding for many reasons as summarized by the Whitesides group[41]:

> " (i) CA is a monomeric, single-chain protein of intermediate molecular weight
> ($\sim$30 kDa), and it has no pendant sugar or phosphate groups and no disulfide
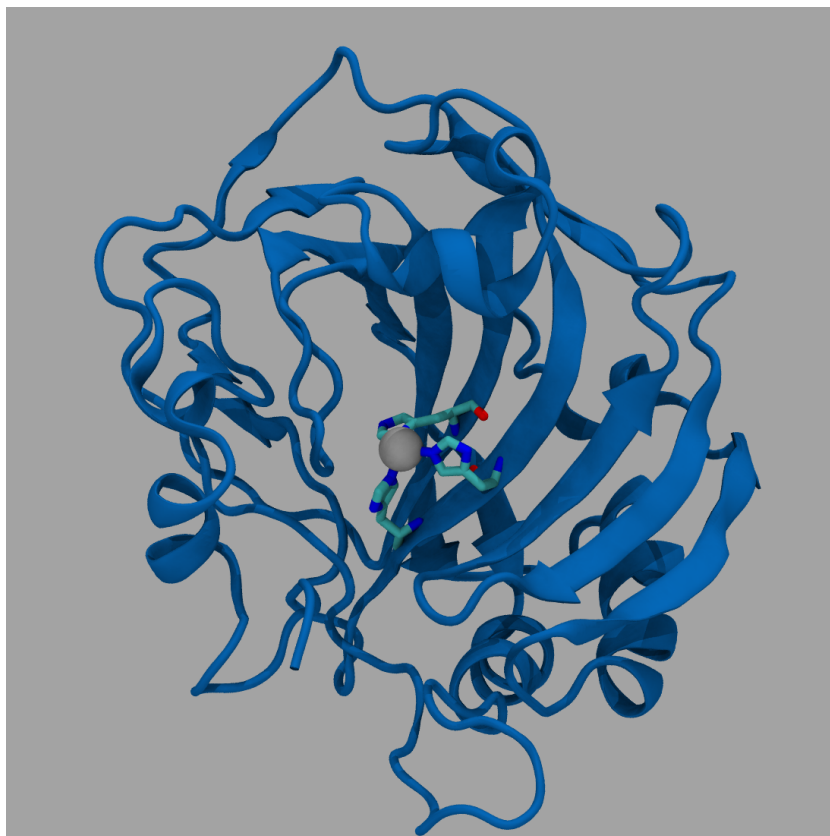
Figure 4.1.: Human carbonic anhydrase 2 features a zinc that is kept in place by three histidines in the center of the protein. The protein is shown as cartoon in blue, the residue are coloured by element and the vdW sphere of zinc in grey.

bonds. (ii) It is inexpensive and widely available. (iii) It is relatively easy to handle and purify, due in large part to its excellent stability under standard laboratory conditions. [...] (vii) The mechanism of inhibition of CA by ligands that bind to the ZnII ion is fairly simple and well-characterized; it is, therefore, easy to screen inhibitors and to examine designed inhibitors that test theories of protein-ligand interactions. [...] ”

4.2. Carbonic Anhydrase as a Host for an Artificial Metalloenzyme

The properties listed by Whitesides *et al.* are very beneficial to an artificial metalloenzyme. The Ward group thus set out to develop hybrid catalysts based on this protein. Arylsulfonamide derivatives are widely studied high affinity inhibitors[158] and thus present themselves as a anchor for the incorporation of a transition metal similar as the biotin in the streptavidin-biotin technology. The first step towards the simulation of such an artificial metalloprotein was the simulation of the protein with several well characterized ligands to validate the approach.

4.3. Arylsulfonamides as inhibitors for carbonic anhydrase: prediction & validation

The zinc metal center contained in this system required a careful evaluation of the possible protein-metal-ligand interaction. Although several zinc treating methods were available, none of them was applicable for (a), maintaining a stable, tetrahedral structure during the simulation, and (b), calculation of the binding free energy of several ligands using MM-GBSA. After exploring several approaches, we employed a custom model and successfully obtained reliable (up to $R = 0.9$) binding free energies for 17 experimentally characterized protein-ligand complexes and predict the influence of several mutations on the binding free energy. The results have been published in Chemical Science.[100]
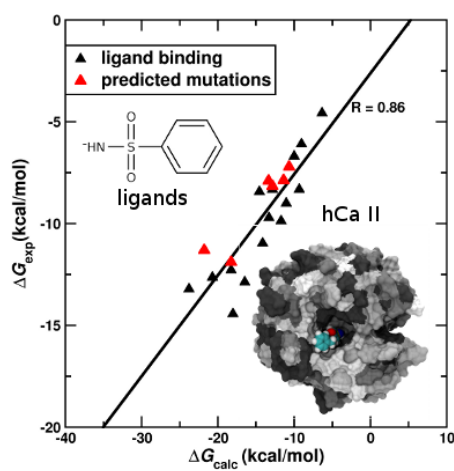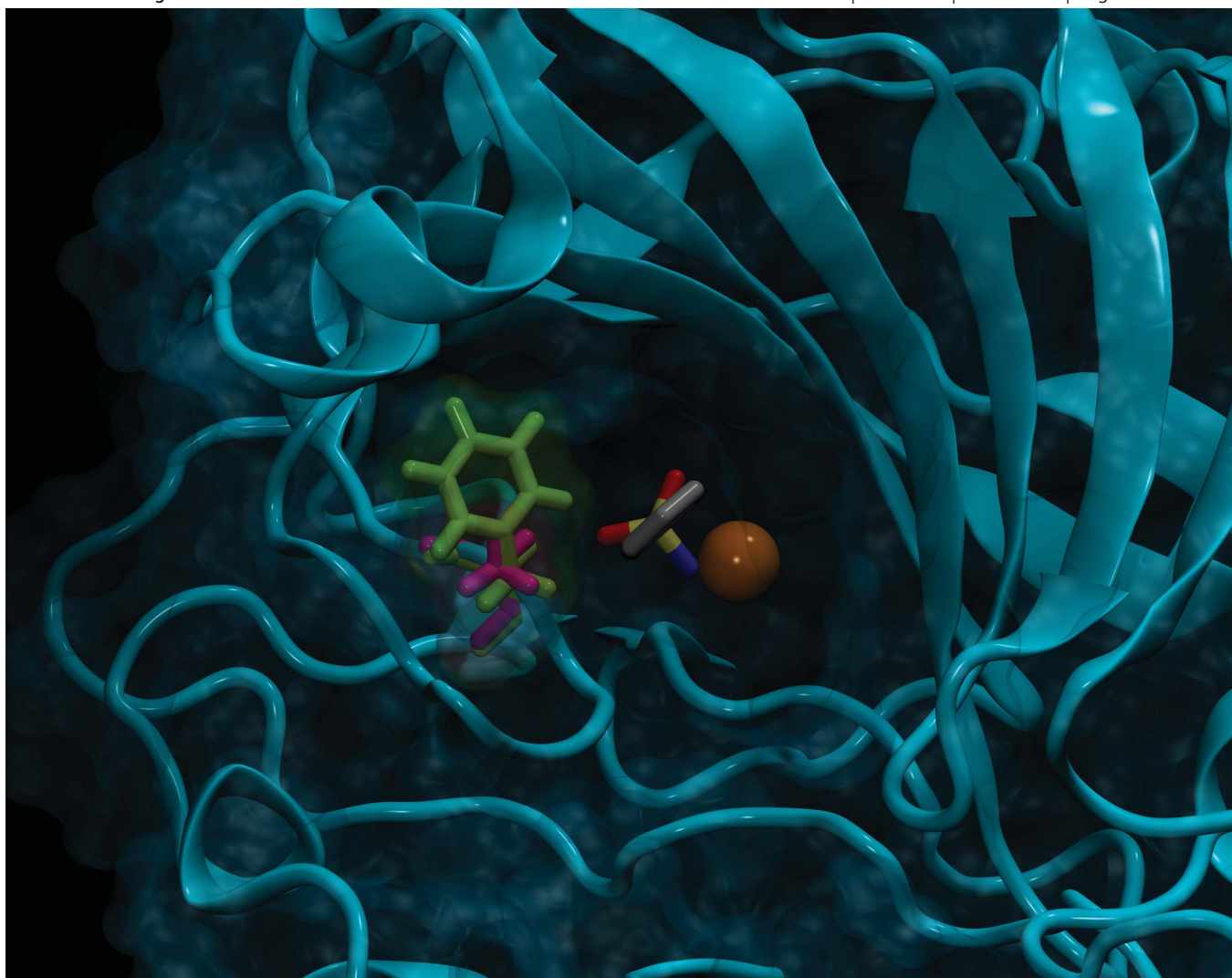
Figure 4.2.: Binding free energies for ligand-protein complexes were determined and the effect of mutations on the binding free energy was predicted. TOC Figure of the article.[100]

Chemical Science

Winner of the ALPSP Award for Best New Journal 2011

RSCPublishing

EDGE ARTICLE
Thomas R. Ward and Markus Meuwly *et al.*
Arylsulfonamides as inhibitors for carbonic anhydrase: prediction & validation

2041-6520(2012)3:3;1-U

59

Chemical Science

Dynamic Article Links ▶

EDGE ARTICLE

Arylsulfonamides as inhibitors for carbonic anhydrase: prediction & validation†

Maurus Schmid,[ab] **Elisa S. Nogueira,**[a] **Fabien W. Monnard,**[a] **Thomas R. Ward**[*a] **and Markus Meuwly**[*b]

Arylsulfonamide derivatives are widely studied high affinity inhibitors of the isozyme human carbonic anhydrase II (hCA II). From molecular dynamics simulations and MM-GBSA calculations, reliable ($R = 0.89$) relative binding free energies are determined for 17 previously experimentally characterized protein–ligand complexes. Decomposition of these energies led to the identification of critical amino acid residues with a significant contribution to the affinity towards the ligands. In particular, Leu198 was predicted as a key residue and was subjected to computational mutagenesis. This prediction was verified experimentally by producing hCA II mutants L198A, L198F and L198Q and determining the resulting affinities towards inhibitor **1**. The computed *vs.* experimental energies are in good agreement thus suggesting that the force field parameters reported herein are useful for the *in silico* design of a wider range of carbonic anhydrase inhibitors.

1 Introduction

Carbonic anhydrases (CA) are ubiquitous metalloenzymes that catalyze the reversible hydration of carbon dioxide with remarkable efficiency ($k_{cat}/K_m \approx 1.5 \times 10^8$ M^{-1} s^{-1}). CA isoforms are involved in various pathological processes including infections, tumorigenicity, osteoporosis, epilepsy, obesity, gluconeogenesis, lipogenesis, ureagenesis, or glaucoma. CAs have thus been the focus of many biophysical studies of protein–ligand interactions. Today, at least 25 clinically used drugs are known to display pronounced CA inhibitory properties.[1]

The active site of most CAs consists of a Zn(His)$_3$ moiety which is essential for catalysis. Thanks to the position of the Zn-cofactor in a cone-shaped cavity (15 Å deep and 15 Å wide at its mouth, see Fig. 1), this system lends itself ideally to the design of potent and selective inhibitors. Among these, arylsulfonamides, which bind tightly to the Zn ion at physiological pH (down to sub-nM), occupy a place of choice.[2]

Determining the binding free energy between proteins and ligands is a formidable and important task in improving ligands or to determine favourable interaction sites of ligands within proteins.[3,4] A range of computational methods are available to address this problem, ranging from free energy perturbation theory and thermodynamic integration to more approximate

procedures, including MM-GBSA. However, routine applications of these methods only has become possible lately and typically retrospective assessments are carried out instead of prospective ones. In the present work we demonstrate that it is possible to predict the effects of protein mutations on ligand binding affinities from atomistic simulation by validating the
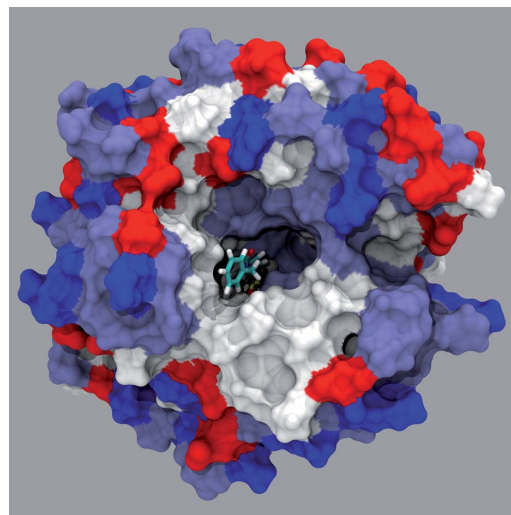

Fig. 1 Rendering of **8** ⊂ hCA II. The ligand is displayed as wireframe and the protein as solvent accessible surface (red: acidic, white: neutral, light blue: polar and dark blue: basic residues).

[a]Department of Chemistry, University of Basel, Spitalstrasse 51, CH-4056 Basel, Switzerland. E-mail: thomas.ward@unibas.ch
[b]Department of Chemistry, University of Basel, Klingelbergstrasse 80, CH-4056 Basel, Switzerland. E-mail: m.meuwly@unibas.ch

computed properties *vis-a-vis* experimentally determined binding constants.

Even though QM/MM models have been applied to metal-containing proteins,[5,6] their computational requirements remain too high for routine usage in the design of metalloenzyme inhibitors. Thus molecular mechanics and empirical force fields are widely used for this purpose. Despite the wealth of structural, kinetic, thermodynamic and quantitative structure–activity relationship data on CA inhibitors,[7] detailed atomistic simulations for a broad range of systems containing Zn–arylsulfonamide binding motifs are rare. Also, a variety of scoring functions have been recently used to score protein–ligand complexes or involving carbonic anhydrase II,[8] including a full QM scoring function.[9] Furthermore, the role of active-site water molecules has been investigated computationally for which dedicated force field parameters were determined (see also supporting information†).[10] The hydrophobic interactions in hCA II have been studied very recently[11] and finally, a computational study of a known sulfonamide inhibitor for two isozymes II and VII of human carbonic anhydrases has been presented.[12]

Because for a wide range of chemically distinct ligands, binding affinities have been determined experimentally (also for selected mutations in the protein), we decided to use hCA II as the system of choice for the present study. The focus in this work is on first validating a simulation strategy for a given set of ligands with experimentally measured ligand binding affinities. For this purpose, we selected MM-GBSA to estimate ligand-binding free energies. To validate this approach, we compared the results with published biophysical data as well as with a simulation using a QM/MM implementation with the Self-consistent charge Density-Functional Tight-Binding (SCCDFTB[13]) method. Finally, having identified amino acid residues critical to binding of arylsulfonamides, hCA II point mutants were computationally investigated and binding free energies were determined. These predictions were compared with experimental biophysical data on the hCA II mutants expressed recombinantly in *E. coli*.

2 Methods

Atomistic simulations were carried out with NAMD[14] and CHARMM.[15] NAMD was used for all simulations with the CHARMM27[16] force field whereas CHARMM was used for analysis and mixed quantum mechanical/molecular mechanics (QM/MM) simulations (see below).

2.1 Molecular dynamics simulations

A specific ligand–protein complex was set up in the following way: When a crystal structure was available, the coordinates from the crystal structure were used for the ligand. When no crystal structure for the ligand ⊂ protein system was available, the minimized energy structure of the ligand (from density functional theory) was docked into the hCA II structure (PDB code 1G54). This structure contains a fluorinated derivative (the five H-atoms of the tailing phenyl are substituted) of **8** (see Fig. 2) as the ligand. To place the new ligand, it was superimposed on the sulfonamide and aryl group of the original ligand present in 1G54 to yield a minimal root mean square deviation (RMSD).
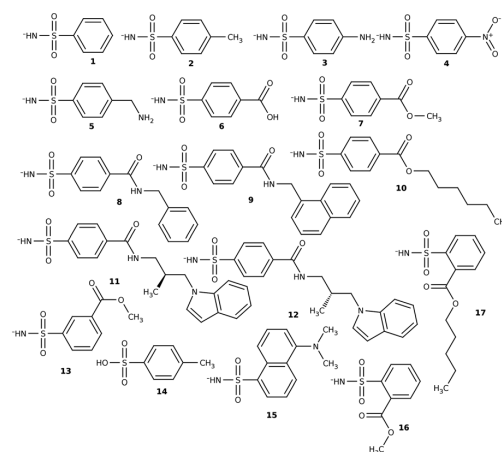

Fig. 2 hCA II inhibitors used in this study.

If the aryl substituents were unsymmetrical (*e.g.* the non *para*-substituted ligand **13**), two alternative structures were generated.

Next, the structure of the protein including the crystal waters and the ligand was minimized in CHARMM for 500 steps using Steepest Descent (SD) to eliminate unfavourable contacts. Then, the system was solvated in a cubic box with dimensions $80 \times 80 \times 80$ Å³ of previously equilibrated water molecules and periodic boundaries were applied. The resulting system consisted of approximately 50000 atoms, depending on the ligand and the number of water molecules added. If the total charge of the system was non-zero due to a charged ligand or charged residues, an adequate number of water molecules was replaced by sodium ions to render the system neutral.

After these setup steps performed in CHARMM, Molecular Dynamics (MD) simulations were performed in NAMD with a time step of $\Delta t = 1$ fs. SHAKE[17] was used on the water molecules. For long-range electrostatic interactions the Particle Mesh Ewald (PME) method was used.[18] Cutoffs for van der Waals (vdW) interactions were 12 Å and a switching scheme was used. Scaled 1–4 parameters were enabled for 1–4 interactions. First, the positions of the water molecules were minimized for 5000 steps of a Conjugate Gradient minimization while keeping the protein and ligand fixed, then the entire system was minimized for the same number of steps. Next, the water molecules were heated to 300 K in increments of 25 K for a total of 12000 steps using a Langevin *NPT* ensemble, keeping the protein and the ligand fixed. In a final step, the entire system was equilibrated for 1 ns in the *NPT* ensemble with all constraints removed and a Langevin dampening coefficient of 5 ps⁻¹. Pressure was controlled with the Nosé–Hoover Langevin piston pressure control. Production runs were typically carried out in blocks of 5 ns with a Langevin dampening coefficient of 1 ps⁻¹.

2.2 QM/MM simulations

To validate the force field parameters, mixed quantum mechanics/molecular mechanics (QM/MM) simulations were carried out using the Self-consistent charge Density-Functional Tight-Binding (SCCDFTB[13]) method implemented in

CHARMM. The QM part for these simulations included the ligand, the zinc atom and the three histidine sidechains bound to the zinc atom. As the sum of the formal charges of the QM atoms is +1, this charge was used for the QM region. Link atoms between the QM- and MM-part were located between the Cα and the Cβ of the histidine residues, see Fig. 3. The system was set up in an analogous fashion as described above. Due to the significantly increased computational cost, only one 8 ns simulation was performed and analyzed.

2.3 Analysis

The binding free energy was calculated using the Molecular Mechanics Generalized Born Surface Area (MM-GBSA[19]) approach. In MM-GBSA the ligand binding free energy ΔG is decomposed in the following way:

$$\Delta G_{bind} = \Delta E_{MM} + \Delta G_{solv} - T\Delta S_{MM} \tag{1}$$

Eqn (1) describes the computation of a binding free energy according to a thermodynamic cycle which includes the enthalpic (E_{MM}) contribution for protein–ligand interactions in the gas phase and the desolvation free energies for the separated and combined protein–ligand complex, respectively. E_{MM} is the "gas phase" energy which is calculated with the CHARMM27[16] force field. E_{MM} contains all internal, electrostatic, and van der Waals energies and the nonbonded interactions are computed without cutoff.

The solvation part ΔG_{solv} consists of a solvent-solvent cavity term (G_{cav}), a solute–solvent van der Waals term (G_{vdW}) and a solute–solvent electrostatic polarization term (G_{pol}), *i.e.*

$$G_{solv} = G_{cav} + G_{vdW} + G_{pol} \tag{2}$$
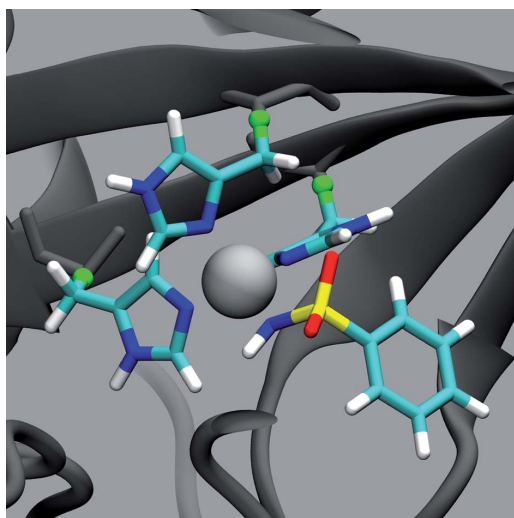
Fig. 3 Close-up view of the QM/MM interface of $\mathbf{1} \subset$ hCA II used for the SCCDFTB simulation. The atoms displayed as coloured wireframe and zinc atom (sphere) were included in the QM part. The dummy atoms (green spheres) form the link between the QM and the MM (black) part of the simulation.

The polarization term in MM-GBSA is calculated within the generalized Born approximation

$$G_{pol} = -166\left(1 - \frac{1}{\varepsilon}\right)\sum\sum\frac{q_i q_j}{f_{GB}} \tag{3}$$

where ε is the dielectric constant of water ($\varepsilon = 80$), q_i and q_j are the charges of atoms i and j. f_{GB} is an expression that depends on the Born atom radii α_i and α_i and distances r_{ij}.[20]

$$f_{GB} = r_{ij}^2 + \alpha_i \alpha_j \exp\left(\frac{-r_{ij}^2}{8\alpha_i \alpha_j}\right) \tag{4}$$

G_{vdW} and G_{cav} are assumed to be linearly related to the solvent accessible surface area (SA):

$$G_{cav} + G_{vdW} = \sum \sigma_k SA_k \tag{5}$$

From the MD simulations snapshots were recorded every 10 ps and the above contributions to ΔG were computed. Then, the energies of these snapshots were averaged to yield final energies and their fluctuations.

The total entropy S consists of translational, rotational and vibrational contributions

$$S_{MM} = S_{trans} + S_{rot} + S_{vib} \tag{6}$$

S_{trans} and S_{rot} depend upon the mass and moments of inertia, whereas calculation of S_{vib} involves normal mode analysis for each frame, which is computationally very costly. It has been previously shown that the influence of S_{vib} on *relative* binding free energies $\Delta\Delta G$ is in generally small and does not affect the ranking of ligands.[21–25] Therefore, contributions to S_{vib} were only calculated every 25 frames. For this, the vibrational entropy for the entire system and for the protein was calculated in CHARMM using the standard normal mode module.

All individual energy contributions ε_i are calculated according to the thermodynamic cycle[22] separately for the entire system, the protein and the ligand. Then the difference $\Delta\varepsilon_i$ between the protein–ligand complex and the separated system is computed for every component:

$$\Delta\varepsilon_i = \varepsilon_i^{Prot-Lig} - (\varepsilon_i^{Prot} + \varepsilon_i^{Lig}) \tag{7}$$

The total binding free energy ΔG can be further decomposed into per-residue contributions ΔG_i to the overall $\Delta G = \sum_i^N \Delta G_i$. Such a decomposition allows us to trace the changes between simulations back to individual residues. This is important in order to identify suitable mutations to enhance or decrease ligand-binding affinities.

Binding free energies from the SCC-DFTB simulations were calculated using the classical force field, as the QM energies are given only as total energies and no decomposition into electrostatic or vdW contributions is possible and the QM part consists of both, the ligand and parts of the protein.

2.4 Ligands examined

The ligands for which relative free energies of binding were determined consisted of sulfonamide ligands displayed in Fig. 2. Most of the ligands consisted of *para*-substituted

arylsulfonamides (**2** to **13**). In addition, two *ortho*-substituted (**16** and **17**) and one *meta*-substituted species were considered. DNSA (**15**) was included in the test set, as well as the sulfonic acid derivative **14**.

To validate the force field, the calculated binding free energies were compared to experimental data. For all ligands included in this study, experimental binding data (K_d or K_i) is available.[2] For some, direct ΔG measurements are available.[2] From K_d, ΔG can be calculated using

$$\Delta G = RTln(K_d). \tag{8}$$

The correlation between experimental and computed (ΔG or $\Delta E_{MM} + \Delta G_{solv}$) binding free energies was considered for both the complete set of ligands as well as the subset containing only the *meta*- and *para*-arylsulfonamides collected in Fig. 2.

2.5 Binding free energies to protein mutants

Binding free energies were computed for the set of ligands and WT hCA II as well as several mutants thereof. Mutants for which experimental data is available include F131V with ligand **8**, for which K_d has been measured.[26] In addition, the calculations suggested (see Results) that L198 contributes significantly to stabilizing the protein–ligand complex. Consequently, mutants L198A, L198F, L198Q were expressed recombinantly and the corresponding ligand binding free energies with ligand **1** were determined. The proteins were mutated *in silico* by substituting the leucine residue with either A, F, or Q, followed by a short optimization using CHARMM. The subsequent simulations and binding free energy calculations were performed as for the native system, described above.

2.6 Recombinant production of hCA II L198X mutants (X = A, F, Q)

Plasmid encoding hCA II and containing a T7 RNA polymerase promoter and an ampicillin resistance gene (pACA)[27] was a generous gift from Prof. Carol Fierke, Michigan University.[28] Ultra competent BL21(DE3)pLysS *E. coli* cells (produced in-house) were transformed by the plasmids containing the desired mutations. Transformed cells were plated on LB-Lennox agar plates containing ampicillin (50 µg mL^{-1}), chloramphenicol (34 µg mL^{-1}) and glucose (2% w/v), and incubated overnight at 37 °C.

One colony was chosen to inoculate 25 mL pre-culture of Luria-Bertoni (LB) medium (10 g L^{-1} tryptone, 5 g L^{-1} yeast extract, 10 g L^{-1} NaCl, 100 µg mL^{-1} ampicillin and 34 µg mL^{-1} chloramphenicol). Pre-culture was grown overnight at 37 °C and 250 rpm. 5 mL of the pre-culture was used to inoculate 1000 mL of induction media (20 g L^{-1} tryptone, 10 g L^{-1} yeast extract, 5 g L^{-1} NaCl, 0.36X M9 salts solution, 0.4% glucose, 60 µM ZnSO$_4$, 100 µg mL^{-1} ampicillin and 34 µg mL^{-1} chloramphenicol). Cells were grown at 37 °C, for 3 to 4 h or until A_{600} = 0.6–0.8. Addition of isopropyl-β-D-thiogalacto-pyranoside (IPTG, 250 µM final concentration) and ZnSO$_4$ (450 µM final concentration) induced protein expression and the temperature was lowered to 18 °C (to prevent formation of inclusion bodies). After overnight incubation at 18 °C the cells were harvested (5346 × g, for 15 min at 4 °C) and frozen at −20 °C overnight. Cells were lysed by activating the gene encoding T7 lysozyme

using three cycles of "freezing/thawing". Cells were resuspended in a buffer containing 50 mM Tris-sulfate, pH 8.0, 50 mM NaCl, 10 mM EDTA, pH 8.0, 0.5 mM ZnSO$_4$, and the protease inhibitor phenylmethanesulfonyl fluoride (PMSF, 10 µg ml^{-1}). Cell resuspension was incubated under vigorous shaking (∼250 rpm) at room temperature for 1 h, DNase I (1 µg L^{-1} final concentration) was added, and cells were left for another hour under the same conditions as previously described. The cellular remnants were centrifuged (12150 × g, for 45 min at 4 °C) and the cell debris was discarded. The supernatant, containing crude hCA II, was purified by affinity chromatography (4-amino-methylbenzene sulfonamide agarose). The column was first equilibrated with 5-column volumes (CVs) of activity buffer (50 mM Tris-sulfate, pH 8.0 and 0.5 mM ZnSO$_4$). The protein was then loaded onto the column. The affinity gel was washed with 5-CVs of 50 mM Na$_2$SO$_4$/50 mM NaClO$_4$/25 mM Tris, pH 8.8. The bound protein was eluted with 10-CVs of 200 mM NaClO$_4$/100 mM NaAc, pH 5.6. Collected fractions were pooled and dialyzed at 4 °C against activity buffer for 24 h, deionized water for another 24 h, and finally against double-deionized (ddH$_2$O) overnight. Protein was lyophilized and kept at 4 °C as a powder for further experiments. This procedure yielded 100–200 mg of >95% pure hCA II, as confirmed by SDS-PAGE. The molecular weight of the three mutants was confirmed by ESI-TOF mass spectroscopy (Bruker micrOTOF II, USA). 0.5 to 1.0 mg of lyophilized protein was dissolved in ddH$_2$O to a final concentration of 1 mg mL^{-1} and further diluted in MS buffer (50% MeOH, 0.05% formic acid, pH 3.0–4.0, and 50% ACN, 0.1% acetic acid and 0.11% TFA, pH 0–1.0). Molecular weight was calculated using Bruker Daltonics DataAnalysis program (Bruker Daltonics, USA). Theoretical and experimental data were compared, and differences were considered not significant (in the range of 0.004 to 0.005%).

2.7 Experimental affinity determination for 1 ⊂ hCA II L198X (X = A, F, Q)

All steady-state measurements[28,29] were performed in Tris-sulfate buffered solution (25 mM, pH 8.0) in presence of 5% DMSO at 25 °C.[30] The organic solvent ensures the solubility of the substrate (*p*-nitrophenyl acetate) as well as of the ligand used as inhibitor. The initial rates of the enzyme-catalyzed activity were measured by following the hydrolysis of the chromogenic substrate, *p*-nitrophenyl acetate, at 348 nm (25 measurements over a period of 35 min). Experiments were carried out in triplicate for each inhibitor. Kinetic measurements were performed in a total reaction volume of 300 µL (in Tris-sulfate buffered solution), containing 0.5 mM *p*-nitrophenyl acetate and different concentrations of inhibitors.

The initial rates of enzyme catalysis were determined using the linear maximum slopes (first 10 min, 5 points) of the reaction traces measured by the plate reader. For comparison of the inhibition data, the initial rates were translated into % activity as a function of the inhibitor concentration (see Fig. 4). The inhibition data were analyzed *via* eqn (9)[29] using gnuplot (Version 4.2, least-squares method)

$$v = \frac{v_o K_i}{K_i + \left\{[I]_t - 0.5\left(A - \sqrt{A^2 - 4[I]_t[E]_t}\right)\right\}} \tag{9}$$

63
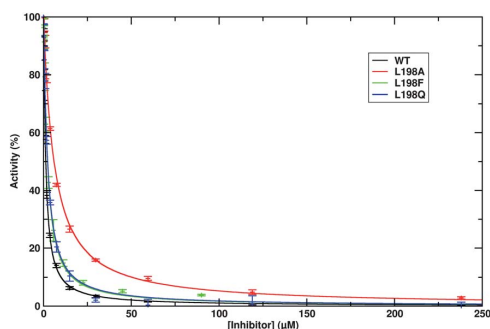
Fig. 4 Steady-state kinetic data for the inhibition of hCA II variants. The solid smooth lines represent the best fits of the data according to eqn (9).

$$A = [I]_t + [E]_t + K_i$$

with v_o being the initial velocity of the enzyme-catalyzed reaction in the absence of inhibitor, K_i the inhibition constant, $[E]_t$ the total concentration of the enzyme and $[I]_t$ the total concentration of the inhibitor.

3   Results and discussion

3.1   Validation of the computational approach

The benzene sulfonamide ligand **1** was selected for an in-depth validation of the computational strategy, in particular of the force field (see supporting information). Various geometrical and energetic properties of the ligand-protein system are considered and analyzed in detail below.

Structural aspects. To assess the structural integrity of the protein–ligand complex, the root mean square deviation (RMSD) along the trajectory relative to the starting structure was considered. For the RMSD analysis, the translational and rotational degrees of freedom of the entire protein were removed by aligning the snapshots to the initial protein structure. This starting structure was also taken as reference for the subsequent RMSD calculations. The RMSD of the coordinates was stable for the entire simulation time which was 32.5 ns. The RMSD of all protein atoms fluctuates around 1.5 Å, while for the backbone and the sidechains of the entire protein, it is around 1 Å and 2 Å, respectively. The RMSD for the ligand itself fluctuates considerably more than the RMSD of the protein atoms. Whereas it ranges from 0.5 Å to 2 Å for the ligand, the protein structure does not fluctuate by more than 0.5 Å.

The backbone atoms of the residues forming the binding pocket are located roughly 7.5 Å from the ligand. As the backbone atoms are stabilized by the secondary structure (see supporting information Figure S3†), they fluctuate less than the sidechain atoms and the RMSD for protein atoms within 7.5 Å of the ligand is lower than the RMSD for protein atoms 5 or 10 Å away. The RMSD was stable for all ligands and simulations with RMSDs in the same range as above.

H-Bonds. Hydrogen bonds can contribute considerably to protein–ligand interactions.[24] Therefore, it is of interest to examine how many H-bonds can and actually are formed between the protein and the ligand and how they are maintained throughout an MD simulation. In the simulation, a total of four distinct hydrogen bonds are observed during the simulation for ligand **1**: The most stable H-bond (as judged from the probability distribution) is the one between the sulfonamide hydrogen H_N and oxygen OG1 of residue Thr199. An additional H-bond can form between the sulfonamide oxygen O_S and hydrogen HN of Thr199. Two further H-bonds can form between HN or HG1 of residue Thr200 and the same oxygen O_S as before (see Fig. 5). The maximum number of H-bonds observed at once between the ligand and the protein is 3. However, this only occurred for short periods of time and the majority of snapshots displays between 0 and 2 H-bonds (see Fig. 6). Occasionally, a water molecule was located between the sulfonamide nitrogen N_S and O_S of residue Thr200. But this occurred only rarely and for short periods of time (≤ 50 ps).

Free enthalpy as a function of time. Binding free energies from MM-GBSA can be calculated either from sufficiently long single trajectories or from several short, independent trajectories.[19,23,31–33] It was previously reported that a minimal simulation time is required to obtain stable, albeit not necessarily converged, binding free energies.[24] In the present work, it was found that during the first nanosecond $<\Delta E_{MM} + \Delta G_{solv}>_{1ns} = -17.6$ kcal mol^{-1}. After 5 ns, this decreased to $<\Delta E_{MM} + \Delta G_{solv}>_{5ns} = -13.8$ kcal mol^{-1} and for the entire trajectory $<\Delta E_{MM} + \Delta G_{solv}>_{32.5ns} = -12.8$ kcal mol^{-1}. Instantaneous values for $\Delta E_{MM} + \Delta G_{solv}$ varied between -40 and $+10$ kcal mol^{-1} with a standard deviation of $\sigma_{\Delta E_{MM} + \Delta G_{solv}} = 7.6$ kcal mol^{-1} (see supporting information†). The largest contributions to $\Delta E_{MM} + \Delta G_{solv}$ arise from G_{pol} and $E_{MM,elstat}$ which approximately compensate each other as $<G_{pol}>$ is positive and $<E_{MM,elstat}>$ is negative.
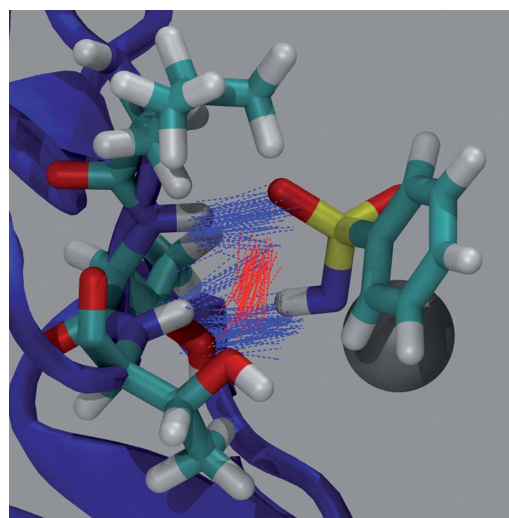
Fig. 5 H-bonding pattern between ligand **1** and hCA II residues Thr199 and Thr200 for 330 snapshots. Red and blue highlight the H-bonds to oxygens and nitrogens respectively.
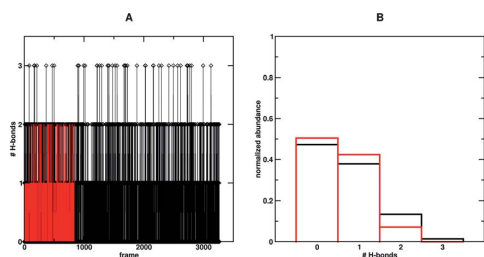
Fig. 6 Evolution of the H-bonding pattern between ligand **1** and protein (**A**) and corresponding histogram (**B**). Values for the MM-Simulation (black), values for the QM/MM simulation (red).

Comparison between MM and QM/MM. The MM simulations (including the force field parametrizations) can be validated to some extent by comparing them with trajectories based on the more elaborate QM/MM approach. The simulations are compared by, *e.g.*, examining geometrical parameters including bond lengths, or by analysing the calculated binding free energy. The distance r_{Zn,N_s} between the zinc and the sulfonamide-nitrogen N_S is compared between the QM/MM and the classical MM simulation. A normalized histogram for the probability $p(r_{Zn,N_s})$ of the bond lengths for all the snapshots displays a slightly wider distribution for the force field simulations compared to QM/MM. The maximum $p_{max}(r_{Zn,N_s})$ is shifted by +0.025 Å in the QM/MM simulations. In the X-ray structure, the distance is $r_{Zn,N_s} = 1.86$ Å, whereas the DFT optimized value for the model complex is 2.05 Å (see Fig. 7). The RMSD of all backbone atoms in the QM/MM simulation is around 1 Å which is comparable to the MM simulation (see supporting information Figure S1†). The RMSD for the ligand fluctuates more in the QM/MM compared to the MM simulation. As above, all RMSDs are reported relative to the starting structure, which is the same for the protein and ligand atoms in MM and QM/MM. The number of hydrogen bonds between the ligand and the protein is, on average, slightly smaller in the QM/MM compared

Fig. 7 Graphical summary of computed Zn-N_S sulfonamide distances for inhibitors **1, 5, 7, 8**.

to the MM simulations: there are at most 2 hydrogen bonds at once, never 3 as in the MM simulation (see Fig. 6).

$<\Delta E_{MM} + \Delta G_{solv}>$ calculated for the trajectory generated with QM/MM was slightly lower (QM = −9.3 *vs.* MM − 12.8 kcal mol^{-1}) than that from the empirical force field. However, the difference is still well within the expected error bars of $\sigma \approx 7$ kcal mol^{-1}.
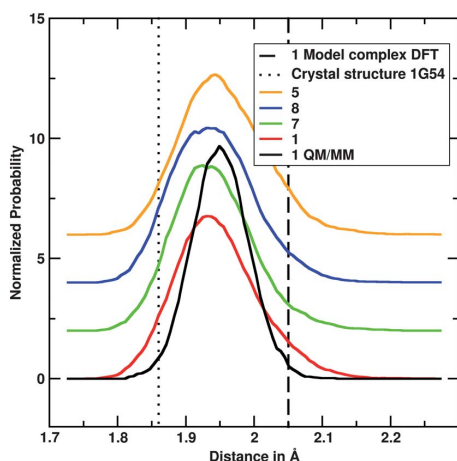
3.2 Simulations for all ligands

After validating the force field and establishing that the computational strategy is robust and meaningful, a broader assessment of the binding free energies for all compounds displayed in Fig. 2 was performed. The primary aim is to follow a simulation strategy which is sufficiently robust to also *predict* the effect of modifications either on the ligand or on the protein on the resulting protein–ligand interaction. Before addressing this last point, a thorough investigation of the 17 compounds is presented. Where necessary, specific ligands are discussed in more detail.

The distribution of the distance between the zinc and nitrogen N_S the ligand is very similar for all ligands except for the sulfonic acid derivative **14**. In this case, the conformationally averaged O–Zn distance is 1.76 Å compared to the 1.95 Å for N_S. For most ligands reported in Fig. 2, a total of 20 to 35 ns were simulated. The results are collected in Table 1. As observed for the model inhibitor **1**, between 10 and 15 ns of simulations are necessary to obtain stable binding free energies. Values for $\Delta E_{MM} + \Delta G_{solv}$ range from −6.4 kcal mol^{-1} for ligand **14** to −23.9 kcal mol^{-1} for ligand **15**. The standard deviations $\sigma_{\Delta E_{MM} + \Delta G_{solv}}$ varied from 6.1 to 8.3 kcal mol^{-1}. The calculated ΔG are between 10.8 kcal mol^{-1} for **14** and −6.4 kcal mol^{-1} for **15**.

For ligand **10**, several calculations with simulation times of 5 ns were performed to investigate the sensitivity to the initial conditions. The calculated binding free enthalpy ranges from −14.6 to −17.9 kcal mol^{-1} with an average of −16.1 kcal mol^{-1}. The ensemble average for a 25 ns simulation is bracketed by these values (−16.5 kcal mol^{-1}). Thus the starting conditions do not greatly influence the simulation results.

For arylsulfonamide ligands with *ortho*-substituents (ligands **16** and **17**), two independent starting structures were generated because no X-ray structure is available. The *ortho*-substituents lead to unstable simulations with significant rearrangement of the ligand position or the surrounding residues. For DNSA (**15**), a commonly used competing ligand for hCA II in fluorescence assays,[34] the crystal structure shows an unusual binding mode with the large aromatic moiety rotated in the hydrophobic subpocket.[35] Two conformations were therefore considered in hCA II: one similar to the X-ray structure (conformation A) and one similar to the other arylsulfonamides, *i.e.* rotated by 180 degrees (conformation B). It was found that conformation A is stabilized relative to conformation B: −23.9 kcal mol^{-1} and −15.2 kcal mol^{-1} for A and B respectively. The hydrophobic interactions within the subpocket appear to be over-estimated compared to other inhibitors where this subpocket is not filled as tightly.

As some of the experimentally measured K_d-values were determined by different groups, a range of values is available for them. In such cases the average value was used to assess the correlation with the computed data. For ligand **1**, the published

65

4. Protein Ligand Simulations

Table 1 Summary of binding data for ligands **1–18** with hCA II

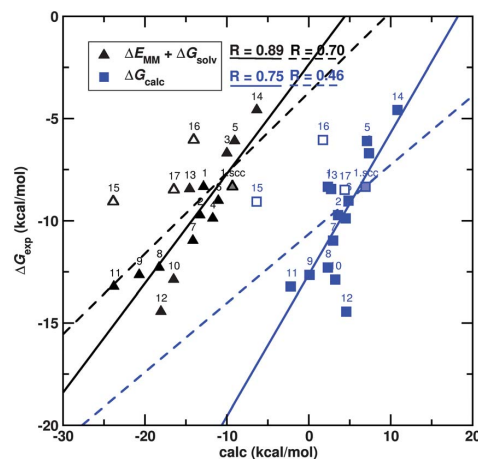|	Calculated					Experimental[a]							
Ligand	ΔE_{MM} + ΔG_{solv} / kcal mol⁻¹	σ	ΔG / kcal mol⁻¹	$-T\Delta S$ / kcal mol⁻¹	t_{sim} /ps	K_d /nM	ΔG_{Kd} / kcal mol⁻¹	ΔG_{obs} / kcal mol⁻¹	ΔH_{obs} / kcal mol⁻¹	$-T\Delta S_{obs}$ / kcal mol⁻¹	$\Delta G_{ArSO_2NH^-}$ / kcal mol⁻¹	$\Delta H_{ArSO_2NH^-}$ / kcal mol⁻¹	$-T\Delta S_{ArSO_2NH^-}$ / kcal mol⁻¹
1	−12.84	±7.64	2.33	15.17	32500	850[b]	−8.33	−9.1	−10.9	1.8	−13.5	−13.1	−0.4
1.scc[c]	−9.33	±8.09	6.93	16.26	8000	850[b]	−8.33						
2[f]	−13.30	±7.98	3.57	13.96	30000	82	−9.72	−9.7	−10.8	1.1	−14.2	−12.1	−2.1
3	−10.01	±6.21	7.35	17.36	35000	13000[b]	−6.70	−6.6	−7.7	1.1	−11.1	−9.3	−1.8
4	−11.72	±6.11	4.94	16.54	25000	63	−9.88	−10.3	−9.5	−0.8	−13.9	−14.2	0.3
5[f]	−9.04	±6.16	7.10	16.14	25000	36000	−6.10	−6.1	−2.4	−3.7	−8.4	−6.5	−1.9
6[f]	−11.05	±6.98	4.85	15.90	25000	270	−9.01	−8.6	−9.6	1.0	−12.6	−10.5	−2.1
7	−14.14	±7.53	2.96	17.09	25000	10	−10.97						
8	−18.27	±6.64	2.33	20.60	25000	1.1	−12.29						
9	−20.71	±6.98	0.09	20.80	25000	0.6	−12.65						
10	−16.49	±6.63	3.22	19.71	25000	0.41	−12.88						
10.1[d]	−17.94	±7.04	0.64	18.58	5000	0.41	−12.88						
10.2[d]	−14.56	±6.51	4.03	18.59	5000	0.41	−12.88						
10.3[d]	−16.85	±5.58	3.92	20.78	5000	0.41	−12.88						
10.4[d]	−15.96	±7.54	2.64	18.60	5000	0.41	−12.88						
10.5[d]	−15.94	±9.28	4.11	20.04	5000	0.41	−12.88						
11	−23.79	±7.04	−2.21	21.60	25000	0.23	−13.22						
12	−18.04	±6.80	4.57	22.61	25000	0.03	−14.44						
13.a	−14.56	±6.78	2.73	17.30	25000	700	−8.44						
13.b[e,h]	−14.04	±6.18	1.88	15.92	5000	700	−8.44						
14	−6.36	±6.36	10.82	17.18	30000	460000	−4.58						
15.a[f]	−23.88	±7.26	−6.38	17.50	25000	250	−9.06	−8.8	−5.7	−3.1	−12.9	−8.2	−4.7
15.b[e,f]	−15.19	±7.75	2.96	18.15	25000	250	−9.06	−8.8	−5.7	−3.1	−12.9	−8.2	−4.7
16.a	−14.04	±8.32	1.72	15.77	20000	39000	−6.05						
16.b[e]	−6.20	±7.16	11.80	18.00	20000	39000	−6.05						
17.a[g]	−16.47	±7.21	4.38	20.85	20000	660	−8.48						
17.b[e,g]	−27.07	±7.62	−6.47	20.60	5000	660	−8.48						

[a] Values from ref. 2. [b] Averaged. [c] SCCDFTB calculation. [d] Different random seed. [e] Alternative starting structure. [f] Experimental decomposed energies from bovine carbonic anhydrase II. [g] Unstable simulation. [h] Becomes the same structure as the other simulation.

K_ds range from 200 to 1500 nM.[2] The average value K_d = 850 nm was used. This translates into a difference in ΔG of ≈ 1 kcal mol⁻¹. In the following, this difference is assumed uniformly as the error in the experiments.

Correlations between (ΔG_{calc}, $\Delta G_{exp,K_d}$) and (ΔE_{MM} + ΔG_{solv}, $\Delta G_{exp,K_d}$) are considered in more detail in the following. This is motivated by the fact that entropic corrections were computed somewhat less rigorously and because previous MM-GBSA studies showed that $-T\Delta S$ had no profound influence on the correlation and did not change the ranking of ligands.[24] Furthermore, experimental data suggests that $-T\Delta S$ is similar for all ligands and therefore leads to a constant offset.[2]

From the correlation between computed and experimentally measured binding free energies (Fig. 8) modest R-values are computed ($R = 0.70$ for ΔE_{MM} + ΔG_{solv} *vs.* $\Delta G_{exp,K_d}$; $R = 0.46$ for ΔG_{calc} *vs.* $\Delta G_{exp,K_d}$). Inspection of the arylsulfonamide structures used in this study reveals three outliers: the *ortho*-substituted ligands **16** and **17** impose steric constraints in the proximity of the Zn-atom which is absent in the *para*- and *meta*-arylsulfonamides **1–14**. Another outlier is the naphtyl derivative **15**. It is interesting to note that the binding free energy and enthalpy calculated for the structure docked in a similar way as the other ligands (conformation B) is much closer to the correlation than the conformation similar to the X-ray structure[35] (conformation A). Without these outliers, correlation increases to $R = 0.89$ (ΔE_{MM} + ΔG_{solv} *vs.* $\Delta G_{exp,K_d}$) and $R = 0.75$ (ΔG_{calc} *vs.* $\Delta G_{exp,K_d}$), respectively.

Additional correlations between ΔG_{calc} and ΔE_{MM} + ΔG_{solv} with ΔG_{obs}, ΔH_{obs}, $\Delta G_{ArSO_2NH^-}$ and $\Delta H_{ArSO_2NH^-}$ were also considered. ΔH_{obs} are the experimentally determined enthalpies

**Fig. 8** Correlation between calculated ΔE_{MM} + ΔG_{solv} (black triangles) and ΔG (blue squares) and experimental ΔG for all ligands (See Fig. 2 for ligand structures). Empty symbols display the outliers and dashed lines the correlation for the outliers. The half-filled symbols (**1.scc**) display the QM/MM results.

This journal is © The Royal Society of Chemistry 2012

66

derived either from temperature dependent ΔG-measurements using the van't Hoff equation $\left(\ln K_d = \dfrac{\Delta H^\circ}{RT} - \dfrac{\Delta S^\circ}{R} \right)$ or using isothermal titration calorimetry. ΔG_{obs} are their associated total binding free energies. $\Delta G_{ArSO_2NH^-}$ and $\Delta H_{ArSO_2NH^-}$ are energies corrected to take into account the deprotonation of the sulfonamide group upon binding to the protein.[2] As all theoretical values are calculated for the deprotonated species, $\Delta G_{ArSO_2NH^-}$ and $\Delta H_{ArSO_2NH^-}$ should correlate better with ΔG_{calc} than ΔG_{obs} and ΔH_{obs}, respectively. It is found that for the experimental enthalpies the R-values remain essentially unchanged, *i.e.* $\Delta E_{MM} + \Delta G_{solv}$ *vs.* ΔH_{obs} and $\Delta H_{ArSO_2NH^-}$ (both $R = 0.84$) and ΔG_{calc} *vs.* ΔH_{obs} and $\Delta H_{ArSO_2NH^-}$ (both $R = 0.83$). However, when considering binding free energies, R-values increase from $R = 0.78$ (ΔG_{calc} *vs.* ΔG_{obs}) to $R = 0.89$ ($\Delta G_{ArSO_2NH^-}$ *vs.* ΔG_{obs}) and from $R = 0.83$ ($\Delta E_{MM} + \Delta G_{solv}$ *vs.* ΔG_{obs}) to $R = 0.89$ ($\Delta E_{MM} + \Delta G_{solv}$ *vs.* $\Delta G_{ArSO_2NH^-}$).

3.3 Per-residue decomposition of $\Delta E_{MM} + \Delta G_{solv}$

It is valuable to trace back the origin of the binding free energy to individual residues. Such information can be used to rationally design mutants for which favourable binding can be perturbed by site directed mutagenesis. Alternatively, the information also provides the basis for tailoring the ligand to the protein to identify which residues play an important role in binding. For the following, refer to Fig. 9 and 10. For all ligands, the dominant stabilizing contribution to the binding free enthalpy can be traced back to residue Leu198. This hydrophobic interaction amounts to almost 4 kcal mol⁻¹. The hydrogen bonds between Thr199 and Thr200 and the sulfonamide moiety add another 2 kcal mol⁻¹ each. Residue His94, which is one of the histidines binding the zinc to the protein, interacts mostly in a nonpolar fashion and provides another 2 kcal mol⁻¹.

For Thr200 there is quite a large difference depending on which ligand is considered. For the majority of ligands, the interaction with Thr200 due to the H-bond is stabilizing but for some of them (ligands **4, 5, 10, 15**) it is either ≈0 or even destabilizing (ligands **12** and **14**). For Phe131 which is located in the mouth of the binding site, the length of the ligand correlates with the magnitude of the influence. Residues Ser29, Arg246 and Arg254 which are not located in the binding site, add 1 kcal mol⁻¹ of stabilization each. Residues Glu106 and Glu117 lead to an
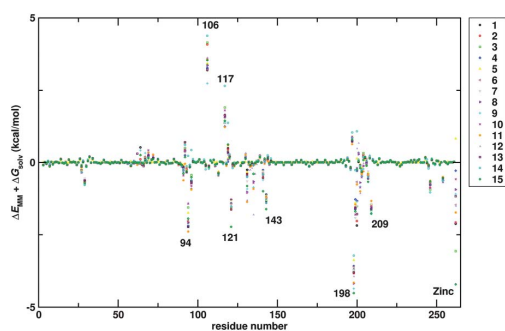
Fig. 9 Per-residue free energies ($\Delta E_{MM} + \Delta G_{solv}$) for ligand ⊂ hCA II combinations.
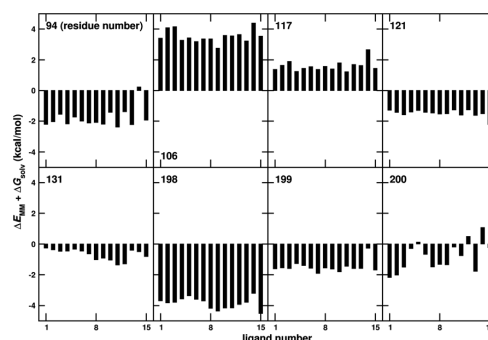
Fig. 10 Free energies ($\Delta E_{MM} + \Delta G_{solv}$) of the ligands (for selected residues see Fig. 2 for ligand structures).

overall destabilization of ≈4 kcal mol⁻¹ and 2 kcal mol⁻¹ respectively.

The influence of the zinc atom changes strongly from one ligand to the other. This effect is related to the distance between the zinc and the sulfonamide nitrogen (see Fig. 7), which, as a result of the other interactions, changes slightly for each ligand. For **14**, with an oxygen binding to the zinc instead of a nitrogen, Glu117 and His119 are much more destabilizing than in the sulfonamides. His94 and Thr199 even lost their stabilizing effect. Especially for **15.a** Val121 plays an important role in stabilization.

In summary, several mutations may be envisaged, especially the residues identified in Fig. 11. For most of the residues identified here, mutation studies have been carried out to investigate either the catalytic function or the binding of the metal cofactor to the protein.[36–38] For F131V, binding constants for arylsulfonamide **8** have been determined.[26]

As the residue which contributes most to the stabilizing interactions is Leu198, we anticipated that mutation at this site may have a significant effect on the resulting affinities. Another candidate would be Glu106, but it is not located directly in the binding site. For Thr200, the difference between ligands are most apparent. Mutations at this position might influence each ligand in a different way. Residue Phe131 is interesting only for ligands with a long tail as it is remote from the binding site.

3.4 Computational and experimental mutation study

Simulation. As experimental data for F131V with ligand **8** is available,[26] it was one of the mutations selected for simulation.
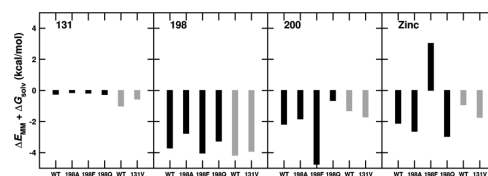
Fig. 11 Contribution to the free energy ($\Delta E_{MM} + \Delta G_{solv}$) for selected residues as a function of ligand (**1** black bars, **8** grey bars) and mutant. (Note the favourable interaction between **1** and residue Thr200 in mutant L198F which is compensated by a strong destabilizing interaction between **1** and **Zn**.)

67

Other mutations concerned Leu198 as this residue has the largest influence on the binding free energy (Fig. 9). The molecular dynamics simulations were performed in a similar way to the ones for the wild-type protein. Again, the RMSDs were stable for all simulations. The number of H-bonds between the protein and the ligand is quite different for L198F and L198Q compared to the WT (see Fig. 12). Whereas there are more H-bonds for L198F (average 0.95) than in the WT (average 0.69), for L198Q (average 0.4) less hydrogen bonding is observed. Fluctuations in $<\Delta E_{MM} + \Delta G_{solv}>$ for the simulations involving the mutants are similar to those for the WT.

Comparing the binding free energy between ligand 1 and WT or the L198F mutant, respectively, a differential stabilization of $\Delta\Delta E_{MM}^{WT\rightarrow L198F} + \Delta\Delta G_{solv}^{WT\rightarrow L198F} = -0.5$ kcal mol^{-1} is found (see Table 2). Considering individual per residue contributions (see Fig. 11), the change in $\Delta\Delta E_{MM} + \Delta\Delta G_{solv}$ at position 198 is essentially 0 whereas Thr200 differentially stabilizes the ligand by more than 2 kcal mol^{-1}. On the other hand, the zinc atom differentially destabilizes the ligand by more than 3 kcal mol^{-1}. No differential changes are found, *e.g.*, at position F131. Therefore, by mutating L198 the contribution of other residues surrounding the active site can be affected in pronounced ways. As will be seen below, this does, however, not affect the correlation between computed and measured ligand binding affinities and suggests, that MM-GBSA simulations are able to capture such effects. When comparing WT with L198Q a differential destabilization of 2 kcal mol^{-1} is found. Again, $\Delta E_{MM} + \Delta G_{solv}$ at position 198 is essentially unchanged, but for Thr200 and the zinc an opposite effect than above can be observed: Thr200

becomes less stabilizing by 2 kcal mol^{-1}, the zinc stabilizes by 1 kcal mol^{-1}. For L198A, which replaces all favourable or unfavourable interactions with the amino acid side chain due to mutation to Ala, the changes for residue 198 are larger than for residue 200. Removing the sidechain from residue 198 leads to subtle changes in the entire binding region and the overall destabilization by ≈ 2 kcal mol^{-1} is the sum of several small contributions. With a valine instead of the phenylalanine at position 131, the stabilization of $8 \subset$ protein for this residue decreases. But already in the wild-type, it is not very large. Finally, its total $\Delta E_{MM} + \Delta G_{solv}$ is larger (-21.8 kcal mol^{-1}) compared to the WT (-18.3 kcal mol^{-1}). The difference cannot be traced back to a single residue but from a sum of several subtle changes.

Experimental dissociation constants. Having identified by computation position L198 as critical in terms of energetic contribution in the affinity of benzenesulfonamide 1 for WT hCA II, three mutants were designed and produced recombinantly in *E. coli*: L198A, L198F and L198Q. The WT hCA II and the three mutants were purified by affinity chromatography and characterized by ESI and SDS-PAGE. Next, the corresponding affinities were determined using the *p*-nitrophenyl acetate hydrolysis assay.[28] This straightforward assay yielded $K_i = 1100 \pm 40$ nM for $1 \subset$ WT hCA II. This value lies well within the published data ranging from 200–1500 nM.[2] The experimentally measured inhibition constants of 1 for the L198X mutants are thus 5500 ± 270 nM for L198A, 1700 ± 130 nM for L198F and 1800 ± 100 nM for L198Q respectively.

Correlation. The experimentally determined binding free energies of the L198X mutants correlate very well with the computed binding free energies (see Fig. 13). Although the differences between the mutants are slightly over-estimated, the calculated energies predict the experimental values quite well. Inclusion of these data into the correlation, affords a correlation coefficient ($R = 0.86$ for $\Delta E_{MM} + \Delta G_{solv}$, *i.e.* 0.03 lower than without the mutants, Fig. 13). As illustrated for the L198X mutants, we have demonstrated that it is possible to correctly predict the influence of mutations at key positions in hCA II from atomistic simulations. Although the influence of the point mutations may, at first, seem subtle compared to the range of the ligands tested, this difference translates into an order of magnitude difference for the corresponding K_ds between the WT hCA II and mutant L198Q.
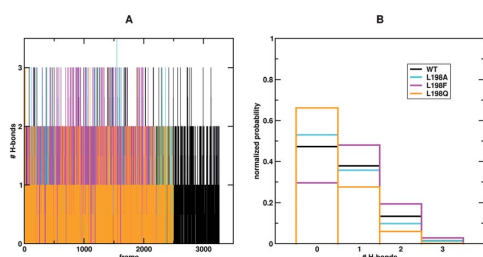
Fig. 12 Evolution of the H-bonding pattern between ligand 1 and hCA II isoform (**A**) and corresponding histogram for WT hCA II and mutants at position Leu198 (**B**).

Table 2 Calculated and experimental thermodynamic data

Ligand	Mutant	Calculated					Experimental	
		$\Delta E_{MM} + \Delta G_{solv}$ /kcal mol^{-1}	σ	ΔG /kcal mol^{-1}	$-T\Delta S$ /kcal mol^{-1}	t_{sim} /ps	K_i /nM	ΔG_{K_i} /kcal mol^{-1}
1	WT	−12.84	±7.64	2.33	15.17	32500	1100	−8.17
1	L198A	−10.67	±7.57	4.76	15.42	25000	5500	−7.22
1	L198F	−13.37	±7.34	2.06	15.42	22000	1700	−7.91
1	L198Q	−11.38	±6.95	4.04	15.42	25000	1800	−7.88
8	WT	−18.27	±6.64	2.33	20.60	25000	2.1[a]	−11.90
8	F131V	−21.80	±7.75	−2.35	19.44	25000	5.6[a]	−11.32

[a] Values from ref. 26.
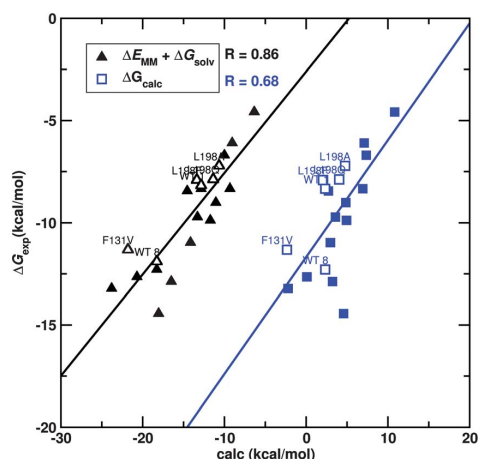
Fig. 13 Correlation between calculated $\Delta E_{MM} + \Delta G_{solv}$ (black triangles) and ΔG (blue squares) and experimental ΔG for WT hCA II (full symbols) and hCA II mutants (empty symbols): $1 \subset L198Q$; $1 \subset L198A$; $1 \subset L198F$; $8 \subset F131V$ (see Fig. 2 for ligand structures).

4   Conclusion

The present work establishes that atomistic simulations for ligand binding in hCA II with validated force fields and sufficiently long conformational sampling allow to reliably rank ligands and predict the effects of mutations on ligand binding affinities. Ligand binding free energies based on MM-GBSA yielded a correlation of $R = 0.89$ calculated *vs.* computed binding free energies between inhibitors **1–14** and hCA II. This corroborates earlier efforts for ligand-binding interactions in HIV-I protease which yielded a correlation of $R = 0.93$.[24] To single out important residues, the binding free energies were decomposed with respect to hCA II individual aminoacids. This led to the identification of Leu198 as a key residue contributing -3.2 to -4.5 kcal mol^{-1} (see Fig. 11) to the interaction between benzenesulfonamide **1** and hCA II. Based on these considerations, three hCA II mutants (L198A, L198F and L198Q) were expressed, purified and tested for their affinity towards benzenesulfonamide **1**. Based on the present results, current efforts are directed towards designing *in silico* specific inhibitors towards carbonic anhydrase isoforms which are overexpressed in certain forms of cancer (*e.g.* hCA IX, hCA XII).

In conclusion, the combined computational and experimental approach to better characterize arylsulfonamide-based ligands interacting with hCA II shows considerable potential to extend the research towards ligand design along similar lines. The experimentally determined binding free energies show good correlation with the predicted energies, thus suggesting that the computational strategy (force field parameters, MD simulations in explicit solvation and of sufficient length, MM-GBSA and decomposition of free energies) presented herein may be widely applicable.

References

1 C. T. Supuran, *Nat. Rev. Drug Discovery*, 2008, **7**, 168–181.
2 V. M. Krishnamurthy, G. K. Kaufman, A. R. Urbach, I. Gitlin, K. L. Gudiksen, D. B. Weibel and G. M. Whitesides, *Chem. Rev.*, 2008, **108**, 946–1051.
3 W. L. Jorgensen, *Science*, 2004, **303**, 1813–1818.
4 W. L. Jorgensen, *Acc. Chem. Res.*, 2009, **42**, 724–733.
5 D. Hartsough and K. Merz, *J. Phys. Chem.*, 1995, **99**, 11266–11275.
6 U. Ryde, *Curr. Opin. Chem. Biol.*, 2003, **7**, 136–142.
7 G. Melagraki, A. Afantitis, H. Sarimveis, O. Igglessi-Markopoulou and C. T. Supuran, *Bioorg. Med. Chem.*, 2006, **14**, 1108–1114.
8 R. Wang, Y. Lu, X. Fang and S. Wang, *J. Chem. Inf. Model.*, 2004, **44**, 2114–2125.
9 K. Raha and K. M. Merz, *J. Am. Chem. Soc.*, 2004, **126**, 1020–1021.
10 D. Lu and G. A. Voth, *Proteins: Struct., Funct., Genet.*, 1998, **33**, 119–134.
11 P. W. Snyder, J. Mecinovic, D. T. Moustakas, S. W. Thomas, M. Harder, E. T. Mack, M. R. Lockett, a. Heroux, W. Sherman and G. M. Whitesides, *PNAS*, 2011, 1–6.
12 J.-L. Zhang, Q.-C. Zheng and H.-X. Zhang, *Comput. Biol. Chem.*, 2011, **35**, 50–56.
13 M. Elstner, D. Porezag, G. Jungnickel, J. Elsner, M. Haugk, T. Frauenheim, S. Suhai and G. Seifert, *Phys. Rev. B: Condens. Matter*, 1998, **58**, 7260–7268.
14 J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kale and K. Schulten, *J. Comput. Chem.*, 2005, **26**, 1781–1802.
15 B. R. Brooks, R. E. Bruccoleri, D. J. Olafson, D. J. States, S. Swaminathan and M. Karplus, *J. Comput. Chem.*, 1983, **4**, 187–217.
16 A. D. MacKerell Jr., C. L. Brooks III, L. Nilsson, B. Roux, Y. Won and M. Karplus, in *CHARMM: The Energy Function and Its Parameterization with an Overview of the Program*, ed. P. v. R. Schleyer *et al.*, John Wiley & Sons: Chichester, 1998, vol. 1, pp. 271–277.
17 J.-P. Ryckaert, G. Ciccotti and H. J. C. Berendsen, *J. Comput. Phys.*, 1977, **23**, 327–341.
18 T. Darden, D. York and L. Pedersen, *J. Chem. Phys.*, 1993, **98**, 10089–10092.
19 W. C. Still, A. Tempczyk, R. C. Hawley and T. Hendrickson, *J. Am. Chem. Soc.*, 1990, **112**, 6127–6129.
20 M. S. Lee, M. Feig, F. R. Salsbury and C. L. Brooks, *J. Comput. Chem.*, 2003, **24**, 1348–1356.
21 W. Wang and P. a. Kollman, *J. Mol. Biol.*, 2000, **303**, 567–582.
22 H. Gohlke, C. Kiel and D. A. Case, *J. Mol. Biol.*, 2003, **330**, 891–913.
23 V. Zoete, M. Meuwly and M. Karplus, *Proteins: Struct., Funct., Bioinf.*, 2005, **61**, 79–93.
24 H. Thorsteinsdottir, T. Schwede, V. Zoete and M. Meuwly, *Proteins: Struct., Funct., Bioinf.*, 2006, **65**, 407–423.
25 T. Hou, J. Wang, Y. Li and W. Wang, *J. Chem. Inf. Model.*, 2011, **51**, 69–82.
26 C.-Y. Kim, J. S. Chang, J. B. Doyon, T. T. Baird, C. A. Fierke, A. Jain and D. W. Christianson, *J. Am. Chem. Soc.*, 2000, **122**, 12125–12134.
27 F. Studier and B. Moffatt, *J. Mol. Biol.*, 1986, **189**, 113–130.
28 S. K. Nair, T. L. Calderone, D. W. Christianson and C. A. Fierke, *J. Biol. Chem.*, 1991, **266**, 17320–17325.
29 D. K. Srivastava, K. M. Jude, A. L. Banerjee, M. Haldar, S. Manokaran, J. Kooren, S. Mallik and D. W. Christianson, *J. Am. Chem. Soc.*, 2007, **129**, 5528–5537.
30 D. Vullo, M. Franchi, E. Gallori, J. Antel, A. Scozzafava and C. Supuran, *J. Med. Chem.*, 2004, **47**, 1272–1279.
31 S. Genheden and U. Ryde, *J. Comput. Chem.*, 2010, **31**, 837–846.
32 H. Gouda, I. Kuntz, D. Case and P. Kollman, *Biopolymers*, 2003, **68**, 16–34.

69

4. Protein Ligand Simulations

33 H. Gohlke and D. A. Case, *J. Comput. Chem.*, 2004, **25**, 238–250.
34 A. L. Banerjee, M. Swanson, B. C. Roy, X. Jia, M. K. Haldar, S. Mallik and D. K. Srivastava, *J. Am. Chem. Soc.*, 2004, **126**, 10875–10883.
35 S. Nair, D. Elbaum and D. Christianson, *J. Biol. Chem.*, 1996, **271**, 1003–1007.
36 S. Nair and D. Christianson, *Biochemistry*, 1993, **32**, 4506–4514.
37 S. K. Nair, J. F. Krebs, D. W. Christianson and C. A. Fierke, *Biochemistry*, 1995, **34**, 3981–3989.
38 S. Z. Fisher, C. Tu, D. Bhatt, L. Govindasamy, M. Agbandje-McKenna, R. McKenna and D. N. Silverman, *Biochemistry*, 2007, **46**, 3803–3813.

70

Supporting information for: Arylsulfonamides as Inhibitors for Carbonic Anhydrase: Prediction & Validation

Maurus Schmid,[†,‡] Fabien W. Monnard,[†] Elisa S. Nogueira,[†] Thomas R. Ward,[†]
and Markus Meuwly[*,‡]

Department of Chemistry, University of Basel, Spitalstrasse 51, CH-4056 Basel, Switzerland, and

Department of Chemistry, University of Basel, Klingelbergstrasse 80, CH-4056 Basel, Switzerland

E-mail: thomas.ward@unibas.ch,m.meuwly@unibas.ch

Contents

[*]To whom correspondence should be addressed
[†]Department of Chemistry, University of Basel
[‡]Department of Chemistry, University of Basel

Force Field Parametrization

Force field parametrization was carried out by comparing with reference electronic structure calculations. First, the structure of the ligand was optimized with Density Functional Theory (DFT[1]) using Gaussian[2] and the hybrid B3LYP functional.[3,4] The atomic basis sets consisted of an effective core potential (LanL2DZ[5]) for the metal and the explicit 6-31G(d,p) basis for all remaining atoms.

All ligands required custom parameters for the force field. They were either taken from parameters for similar models through analogies or derived by fitting CHARMM force field terms[6] to reproduce energy surface scans of the respective parameters calculated with DFT, as described above. To assist the parametrization, perl scripts using PerlMol[7] were used to create the topology and necessary input files for the CHARMM calculations directly from the Gaussian output files. The fitting was performed using CHnolls[8], the CHARMM interface for Inolls[9]. The charges for the ligand atoms were set to the Mulliken charges from the DFT optimizations (see above).

Two approaches can be envisaged for parametrizing the $Zn(His)_3$ centre in CAs: A "non–bonded"[10] and a "bonded" one[11–13]. Recently, a systematic derivation of force field parameters for the bonded model was reported by Lin et al.[14] In the context of MM-GBSA, however, the fully bonded approach is not suitable because calculating a ligand-binding free energy ΔG requires the protein and the ligand to be non-covalently bound entities. Furthermore, test simulations with the fully non-bonded model failed to reproduce the tetrahedral geometry of the zinc centre. To circumvent this, a cationic dummy atom model was developed by Pang.[15]

Inspired by these previous efforts, a model consisting of a combination of bonded and non–bonded terms, was envisaged: The zinc center was parametrized with explicit bonds to the protein, but devoid of a bond to the arylsulfonamide ligand. This is motivated by the fact that the computation of ligand binding free energies (see below) is more straightforward with such a model. The zinc atom thus had a bond to each of the three histidine nitrogen atoms $His_{94}N_\varepsilon$, $His_{96}N_\varepsilon$, and $His_{119}N_\delta$, but none to the nitrogen atom N_S of the sulfonamide. Some of the parameters for the zinc-histidine interactions from Lu and Voth[12], who parametrized an approach with four bonds to the zinc, proved

to be valid here as well and were used subsequently. The charge of the zinc atom was set to $q_{Zn} = +1e$, which was supported by a DFT calculation for a model system of the binding site consisting of 3 imidazoles and a arylsulfonamide, where the resulting Mulliken charge on zinc was $+1.04\ e$. The parameters for the zinc center are given in Table S2.

To maintain a tetrahedrally coordinated zinc ion, as observed in the crystal structures, constraints on the zinc binding site were introduced using the Colvars module of NAMD. The 3 valence angles $N_S ZnN_{\varepsilon,\alpha}$, where $\alpha = 1, 2, 3$ for the three His-nitrogen atoms, were constrained with a harmonic constraint $k_\theta (\theta - \theta_0)^2$ with $\theta_0 = 120°$ with a rescaled force constant of 0.05 kcal/mol/deg^2. Further harmonic potentials were set on the 3 dihedral angles formed by these atoms and the sulphur S of the ligand. The dihedrals were constrained to the corresponding values observed in the crystal structure with a rescaled force constant of 0.00111 kcal/mol/deg^2. These values proved to yield a stable and conserved tetrahedral structure for the binding site throughout all simulations. The constraints were only used during the simulations and not for the subsequent analysis.

General Aspects

Materials and reagents were purchased from the highest commercially available grade and used without further purification. Esterase activity was recorded on a Tecan Safire spectrophotometer using NUNC 96-well plates. Human Carbonic Anhydrase isozyme II (hCA II) was expressed in 1 L shaking-flasks in an Infors HT Ecotron shaker and culture growth was checked by UV-Vis at 600 nm with a Varian Cary 50 Scan. Cell cultures were centrifuged either using a Heraeus Multifuge 4KR or a Heraeus Suprafuge 22 (5346 x *g*). Cells were resuspended using Lab-Shaker (Adolf Kükner AG, Switzerland). SDS-PAGE gels were analyzed on a BioRad Gel Doc XR (Software: Quantity One 4.6.6). Sulfonamide affinity chromatography was performed using ÄKTAprime (Amersham Pharmacia Biotech, Software: PrimeView 5.0). Molecular weight was confirmed by MS ES-TOF (Bruker micrOTOF II, USA) and analyzed with Bruker Daltonics Data-Anakysis software.

Site-directed mutagenesis

Plasmid encoding hCA II and containing a T7 RNA polymerase promoter and an ampicillin resistance gene (pACA)[16] was a generous gift from Carol Fierke, Michigan University.[17] The construct of this plasmid has a serine residue at position 2 instead of an alanine, with no effect on protein expression or catalytic properties.

Primers were designed following the method described by Zheng et al.[18] and tested *in silico* to minimize hairpin formation (Kibbe, www.basic.northwestern.edu/biotools/OligoCalc.html). Primers were obtained from Microsynth Balgach, Switzerland. PCR reactions were prepared by addition of 5 μL 10x Pfu buffer, 2 μL of 10 mM dNTP (final concentration 0.4 mM), 2.5 μL DMSO (final concentration 5 %), 1.5 μL Pfu Turbo polymerase, 1.5 μL of 10 μM primers (forward and reverse), 35 μL H_2O to 1 μL of template. The cycle conditions were: initial denaturation (95 °C, 5 min), followed by 16 cycles of 1 min at 95 °C; 1 min at 60 °C; 15 min at 68 °C. The final elongation was performed at 68 °C for 1 h. PCR products were analyzed by 2.4 % agarose gel electrophoresis.

The initial DNA template (wild type sequence) was digested by *Dpn*I (4 h at 42 °C). 5 μL of PCR product was used to transform ultra competent XL1-blue *E. coli* cells (produced in-house). Plasmids were purified using a Wizard Plus SV Miniprep DNA purification System (Promega - Switzerland) and were sequenced either by Starseq (Germany) or Microsynth (Switzerland).

Tables

S4

74

Table S1: Primers sequences for Q-PCR

Mutant	Forward primer (+)	Reverse primer (-)	Melting temp (°C)	Product length
L198A	CCTACCCAGGCT CAGCGACCACCC CTCCTCTTCTG	GAGGAGGGGTGG TCGCTGAGCCTG GGTAGGTCC	72.7(+)/ 73.5(-)	35(+)/33(-)
L198F	CCTACCCAGGCT CATTTACCACCC CTCCTCTTCTG	GAGGAGGGGTGG TAAATGAGCCTG GGTAGGTCC	70.2(+)/ 71.0(-)	35(+)/33(-)
L198Q	CCTACCCAGGCT CACAGACCACCC CTCCTCTTCTG	GAGGAGGGGTGG TCTGTGAGCCTG GGTAGGTCC	71.5(+)/ 71.9(-)	35(+)/33(-)

Table S2: MM Parameters for the Zinc moiety used in this study

Atoms				Force constant		equilibrium
Bonds				k		r_e
Zn	NR2			40 kcal/mol/Å^2		2.24 Å
Angles[a]				k		θ_e
NR2	Zn	NR2		23 kcal/mol/rad^2		109.5°
CPH2	NR2	ZNB		20 kcal/mol/rad^2		126.0°
Dihedrals				k	n	φ_0
NR1	CPH2	NR2	ZNB	5.0 kcal/mol	2	180.0°
CPH1	CPH1	NR2	ZNB	5.0 kcal/mol	2	180.0°
HR1	CPH2	NR2	ZNB	5.0 kcal/mol	2	180.0°
HR3	PH1	NR2	ZNB	5.0 kcal/mol	2	180.0°
Improper Dihedrals				k	n	φ_0
ZNB	NR2	NR2	NR2	25 kcal/mol/rad^2	0	0°

[a] parameters from [12]

S5

Supplementary Figures

Figure captions

Figure S1 Variation of RMSD as a function of simulation time for **1** $\subset$ hCA II.

Figure S2 MM-GBSA interaction energies of **1** $\subset$ hCA II per frame with running averages over 100 steps.

Figure S3 Selected active site residues of **1** $\subset$ hCA II. Residues colored by residue type (red: acidic, white: neutral, light blue: polar and dark blue basic residues).

S6
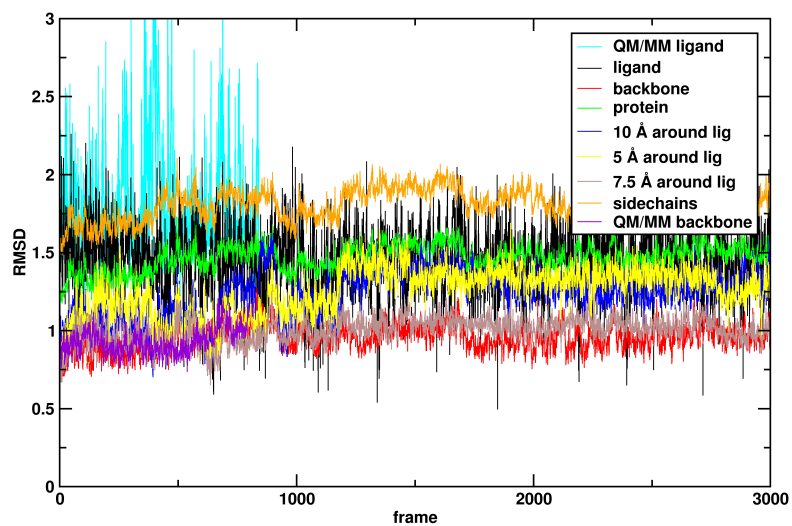
Figures

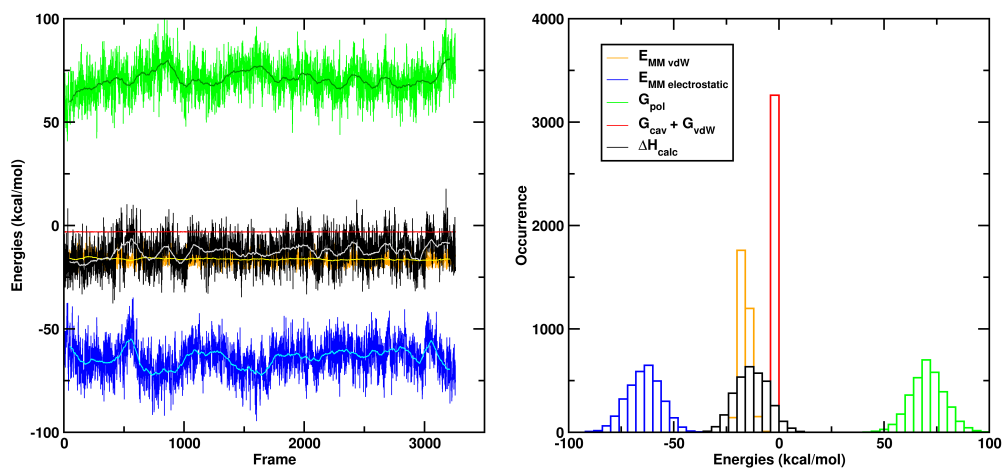Figure S1: Variation of RMSD as a function of simulation time for **1** ⊂ hCA II.

Figure S2: MM-GBSA interaction energies of **1** ⊂ hCA II per frame with running averages over 100 steps.
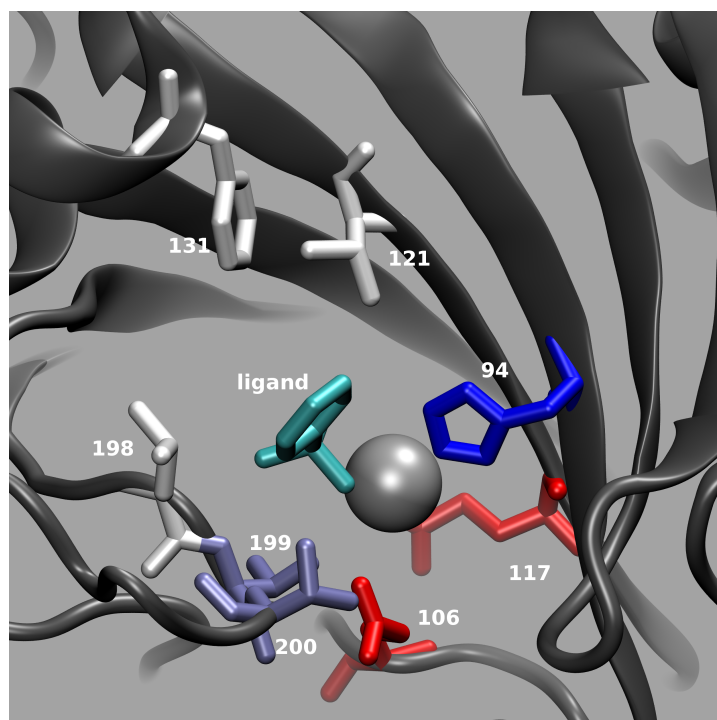

Figure S3: Selected active site residues of **1** ⊂ hCA II. Residues colored by residue type (red: acidic, white: neutral, light blue: polar and dark blue basic residues).

S8

References

(1) Hohenberg, P.; Kohn, W. *Phys. Rev.* **1964**, *136*, B864–B871.

(2) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery Jr., J. A.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03, Revision B.01*.

(3) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648–5652.

(4) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. *J. Phys. Chem.* **1994**, *98*, 11623–11627.

(5) Hay, P. J.; Wadt, W. R. *J. Chem. Phys* **1985**, *82*, 270–283.

(6) MacKerell Jr., A. D.; Brooks III, C. L.; Nilsson, L.; Roux, B.; Won, Y.; Karplus, M. In *CHARMM: The Energy Function and Its Parameterization with an Overview of the Program*; v. R. Schleyer et al., P., Ed.; John Wiley & Sons: Chichester, 1998; Vol. 1, pp 271–277.

(7) Tubert-Brohman, I. *The Perl Journal* **2004**, *8*, 3–5.

(8) Devereux, M.; Meuwly, M. *J. Chem. Inf. Model.* **2010**, *50*, 349–357.

S9

(9) Law, M.; Hutson, J. *Comput. Phys. Commun.* **1997**, *102*, 252–268.

(10) Stote, R. H.; Karplus, M. *Proteins Struct. Funct. Bioinf.* **1995**, *23*, 12–31.

(11) Hoops, S. C.; Anderson, K. W.; Merz, K. M. *J. Am. Chem. Soc.* **1991**, *113*, 8262–8270.

(12) Lu, D.; Voth, G. A. *Proteins Struct. Funct. Bioinf.* **1998**, *33*, 119–134.

(13) Alterio, V.; Vitale, R. M.; Monti, S. M.; Pedone, C.; Scozzafava, A.; Cecchi, A.; De Simone, G.; Supuran, C. T. *J. Am. Chem. Soc.* **2006**, *128*, 8329–8335.

(14) Lin, F.; Wang, R. *J. Chem. Theory Comput.* **2010**, *6*, 1852–1870.

(15) Pang, Y.-P. *J. Mol. Model.* **1999**, *5*, 196–202.

(16) Studier, F.; Moffatt, B. *J. Mol. Biol.* **1986**, *189*, 113–130.

(17) Nair, S. K.; Calderone, T. L.; Christianson, D. W.; Fierke, C. A. *J. Biol. Chem.* **1991**, *266*, 17320–17325.

(18) Zheng, L.; Baumann, U.; Reymond, J.-L. *Nucleic Acids Res.* **2004**, *32*, e115.

5. Modelling of Metal Complexes in Strepativdin

The usefulness of calculations for a rational understanding of the binding mode of metal complexes in artificial metalloenzymes can be illustrated by the following study. To yield a more controlled metal location by establishing a dual anchored metal, docking studies were performed. Introduction of the metal catalyst into streptavidin is mainly performed using supramolecular anchoring. A drawback of this method is the fact that the metal is not anchored by itself, but by a linker group. The metal position is not directly defined but a result of linker length, cavity space and interactions with surrounding residues. In fact, the metal may be oscillating between different stable positions and not be exclusively fixed at one position. Adding a coordinative bond between the protein and the metal increases its positional stability. To achieve this, two interacting conditions need to be fulfilled: (a) An appropriate residue needs to be chosen for mutation to a residue capable of metal binding (like histidine), and (b) the spacer between the metal and the biotin anchor needs to have the right length to bring the metal close to this residue. Docking with several spacers was employed to test the possibility of a metal bond, short MD simulations were performed to yield docked structures. Although a crude simulation scheme (purely CHARMM force field with NAMD[159] and part of the system fixed) was employed, this valuable information helped developing and understanding the system. A manuscript has been published in the Journal of the American Chemical Society.[160]

5.1. A dual anchoring strategy for the localization and activation of artificial metalloenzymes based on the biotin-streptavidin technology

A Dual Anchoring Strategy for the Localization and Activation of Artificial Metalloenzymes Based on the Biotin−Streptavidin Technology

Jeremy M. Zimbron, Tillmann Heinisch, Maurus Schmid, Didier Hamels, Elisa S. Nogueira, Tilman Schirmer, and Thomas R. Ward*

Biozentrum and Department of Chemistry, University of Basel, CH-4056 Basel, Switzerland

Ⓢ *Supporting Information*

ABSTRACT: Artificial metalloenzymes result from anchoring an active catalyst within a protein environment. Toward this goal, various localization strategies have been pursued: covalent, supramolecular, or dative anchoring. Herein we show that introduction of a suitably positioned histidine residue contributes to firmly anchor, via a dative bond, a biotinylated rhodium piano stool complex within streptavidin. The in silico design of the artificial metalloenzyme was confirmed by X-ray crystallography. The resulting artificial metalloenzyme displays significantly improved catalytic performance, both in terms of activity and selectivity in the transfer hydrogenation of imines. Depending on the position of the histidine residue, both enantiomers of the salsolidine product can be obtained.
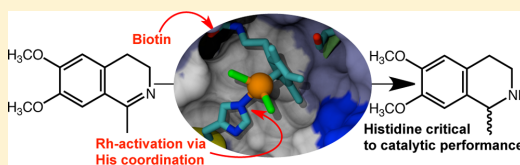
INTRODUCTION

In recent years, the field of artificial metalloenzymes has attracted increasing attention as an alternative to the more traditional heterogeneous, homogeneous, and enzymatic catalysts.[1−6] Artificial metalloenzymes result from incorporation of a catalytically competent organometallic moiety within a protein scaffold. Toward this goal, covalent, dative, and supramolecular anchoring strategies have been pursued.

We and others have been relying on the biotin−(strept)-avidin technology to create artificial metalloenzymes.[7−11] The catalytic performance of these can be optimized relying on either chemical or genetic means.[12] Recent structural and kinetic data of artificial metalloenzymes based on the biotin−streptavidin technology highlighted that the metal is located in a shallow energy minimum resulting in poorly localized metal moieties (as reflected by low occupancy in the X-ray structure) and that the presence of a second coordination sphere provided by streptavidin significantly influences the selectivity of the resulting artificial metalloenzymes.[13] In contrast, the activity (TOF, TON) is only moderately affected upon incorporation in streptavidin. This presents a serious limitation toward the implementation of directed evolution protocols of artificial metalloenzymes. Indeed, it requires the precise quantification of the protein as any nonprotein-bound metal would lead to an erosion of enantioselectivity.[10] We speculated that, in addition to the biotin anchor, a properly positioned histidine residue provided by streptavidin may contribute to firmly localize the metal-containing biotinylated cofactor and to activate the precatalyst (Scheme 1). The groups of Lu and Watanabe independently pioneered an anchoring strategy to introduce Schiff-base complexes within apo-myoglobin. In both strategies,

the Schiff-base complex was anchored thanks to a dative bond with the proximal histidine 93.[14,15] In addition, Lu and co-workers introduced two cysteine residues that further contributed to firmly localize the artificial cofactor thanks to the formation of two disulfide bonds with the Schiff-base.[14] Compared to the artificial metalloenzyme lacking the disulfide bridges, the dually anchored artificial metalloenzymes displayed improved sulfoxidase activity.[14] Inspired by this elegant study, we present our efforts toward the design and application of a dually anchored (i.e., dative and supramolecular) artificial metalloenzyme for the reduction of prochiral imines.

RESULTS AND DISCUSSION

Design Considerations. With the aim of generating a widely applicable d[6]-piano stool complex for a variety of catalytic applications upon introduction in streptavidin (Sav hereafter), we set out to link the biotin anchor to the cyclopentadienyl moiety. This strategy leaves three coordination sites available for catalysis and/or activation via additional ligands (Scheme 1). In order to gain a semiquantitative insight on the position of the biotinylated piano stool complex **5** upon incorporation within Sav and which amino acid position may be amenable to bind to the metal upon mutation to a histidine, docking simulations were performed. All X-ray crystal structures containing a biotinylated metal complex obtained within the Ward group indicate that the position of the biotin moiety and the C_αs of the protein is essentially invariant, irrespective of the nature of the metal complex bound to

Scheme 1. Dual Anchoring Strategy to Localize and Activate a Biotinylated Piano Stool Catalyst Precursor upon Incorporation into Sav
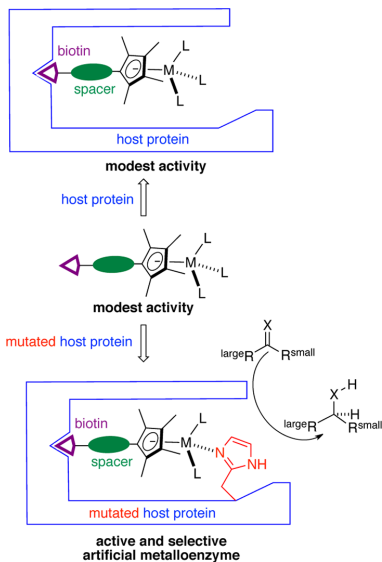
biotin.[13,16–18] Therefore, the DFT-optimized structure of the monomeric biotinylated complex $[\eta^5\text{-(Biot-}2)\text{RhCl}_2(\text{H}_2\text{O})]$ was manually positioned into WT Sav (PDB code 1MK5) by superimposing the biotin anchor of the complex to the native biotin present in the crystal structure. Simulations were performed in NAMD[19] using the CHARMM27 force-field[20] with parameters developed in our group[21] or elsewhere[22] (see SI for details).

The system was subsequently minimized for 1000 conjugate gradient steps, followed by a short molecular dynamics simulation (Langevin NVT ensemble) of 500 ps. The structure in the final snapshot was minimized again for 1000 steps to yield the docked structure. Only the aquo complex $[\eta^5\text{-(Biot-}2)\text{RhCl}_2(\text{H}_2\text{O})]$ and atoms within 5 Å of the complex were allowed to relax during the simulation, the rest of the protein was kept frozen. The resulting structure is depicted in Figure 1a. With respect to the rhodium center, the two closest lying C_β atoms are provided by S112 (5.62 Å) and K121 (6.09 Å of the
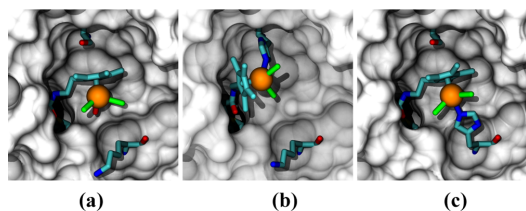
Figure 1. Docked structure of $[\eta^5\text{-(Biot-}2)\text{RhCl}_2(\text{H}_2\text{O})] \subset$ WT Sav (a), $[(\eta^5\text{-Biot-}2)\text{RhCl}_2]$ **5** $\subset$ S112H Sav (b), and $[(\eta^5\text{-Biot-}2)\text{RhCl}_2]$ **5** $\subset$ K121H Sav (the His is provided by the adjacent Sav monomer of the homotetrameric structure) (c). The piano stool complex and Sav's side chains at positions 112 and 121 are displayed as stick representation; Sav as solvent accessible surface.

adjacent monomer). Next, a histidine residue was introduced at either position S112 or K121. The rhodium-bound water was replaced by a bond to the N_ϵ histidine, and the procedure described above was applied yielding $[(\eta^5\text{-Biot-}2)\text{RhCl}_2]$ **5** $\subset$ S112H Sav and $[(\eta^5\text{-Biot-}2)\text{RhCl}_2]$ **5** $\subset$ K121H Sav (Figure 1b,c respectively).

These simulations thus suggest that an ethylene spacer between the Cp* and the biotin moieties positions the metal center in an environment, suitable for additional dative anchoring to a histidine side chain upon mutation of either S112H or K121H. Similar docking simulations were performed for complexes bearing either a methylene or a propylene spacer in the presence of WT Sav. In both cases however, inspection of the docked structure revealed a less propitious environment for dative anchoring to an amino acid side chain of Sav. For the methylene spacer, steric hindrance between the docked ligand and the protein are observed, whereas the propylene spacer projects the metal out of the biotin-binding vestibule (see SI).

Synthesis. Having identified a suitable spacer between the biotin anchor and the Cp* moiety, the corresponding d[6]-piano stool complexes were synthesized. For this purpose, tetramethylcyclopentenone **1** was used as starting material for the synthesis of the cyclopentadienyl precursor H-**2** according to a slightly modified protocol of Teuben and co-workers.[23] Reaction with $\text{RhCl}_3 \cdot (\text{H}_2\text{O})_n$ or $[\text{Ir(COD)Cl}]_2$, followed by an *in situ* oxidation with HCl, yielded the dinuclear piano stool complexes $[\eta^5\text{-(2)}\text{RhCl}_2]_2^{2+}$ **3** and $[\eta^5\text{-(2)}\text{IrCl}_2]_2^{2+}$ **4**, respectively.[24,25] Subsequent biotinylation using the pentafluorophenylester of biotin allowed the isolation of the analytically pure, dimeric $[\eta^5\text{-(Biot-}2)\text{RhCl}_2]_2$ **5**$_2$ and $[\eta^5\text{-(Biot-}2)\text{IrCl}_2]_2$ **6**$_2$, respectively (see SI for details).

Catalysis. As model reaction, we selected the asymmetric transfer hydrogenation (ATH) of the prochiral imine **7** using biotinylated dimeric catalyst precursors **5**$_2$ and **6**$_2$. This reaction is frequently used as a benchmark as it offers a straightforward access to the alkaloid salsolidine **8** (Scheme 2b).[26] Selected

Scheme 2. Synthesis of Biotinylated Piano Stool Complexes (a) and Asymmetric Transfer Hydrogenation Conditions Used Towards the Production of Salsolidine 8 (b)
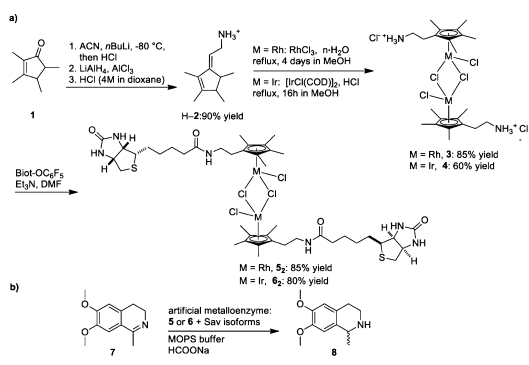
results are collected in Table 1 and can be summarized as follows:

(1) The iridium catalyst precursor $[\eta^5\text{-(Biot-}2)\text{IrCl}_2]_2$ **6**$_2$ leads to higher conversions than the rhodium congener $[\eta^5\text{-(Biot-}2)\text{RhCl}_2]_2$ **5**$_2$. The same trend applies upon incorporation in WT Sav albeit with reduced conversions

Table 1. Selected Results for the ATH of Prochiral Imine 7 Catalyzed by Artificial Metalloenzymes

entry	$[(\eta^5\text{-Biot-L})MCl_2]_2$	Sav isoform	temp (°C)	pH	M loading (%)	conv. (%)	ee (%)
1	Rh	–	55	6.5[a]	2	26	rac
2	Ir	–	55	6.5[a]	2	87	rac
3	Rh	WT	55	6.5[a]	2	18	rac
4	Ir	WT	55	6.5[a]	2	60	6 (S)
5	Rh	S112H	55	6.5[a]	2	95	41 (S)
6	Rh	S112H	55	5.0[b]	1	quant.	55 (S)
7	Rh	S112H	25	5.0[c]	2	85	52 (S)
8	Rh	S112H	5	5.0[b]	1	5	45 (S)
9	Ir	S112H	55	6.5[a]	2	27	11 (S)
10	Rh	K121H	55	6.5[a]	2	95	50 (R)
11	Rh	K121H	55	5.0[b]	1	quant.	79 (R)
12	Rh	K121H	25	5.0[c]	2	25	50 (R)
13	Rh	K121H	5	5.0[b]	1	5	50 (R)
14	Ir	K121H	55	6.5[a]	2	37	rac
15	Rh	S112C	55	6.5[a]	2	93	14 (S)
16	Rh	S112E	55	6.5[a]	2	26	rac
17	Rh	S112K	55	6.5[a]	2	40	13 (S)
18	Rh	S112M	55	6.5[a]	2	98	19 (R)
19	Rh	K121N	55	6.5[a]	2	44	rac
20	Rh	S112H-121H	55	6.5[a]	2	40	9 (R)

[a]The reaction was carried out in a MOPS buffer at pH 6.5 using HCOONa as hydride source: 2 mol % monomeric biotinylated catalyst **5** or **6** (458 μM final metal concentration), 4 mol % biotin binding sites (916 μM final concentration), and substrate **7** (22.9 mM final concentration, total reaction volume 200 μL) (see SI). [b]The reaction was carried out in a MOPS buffer at pH 5.0 using HCOONa as hydride source: 1 mol % monomeric biotinylated catalyst **5** or **6** (458 μM final metal concentration), 2 mol % biotin binding sites (916 μM final concentration), and substrate **7** (45.8 mM final concentration, total reaction volume 200 μL) (see SI). [c]The reactions were carried out in a MOPS buffer at pH 5.0 using HCOONa as hydride source: 2 mol % monomeric biotinylated complex **5** (680 μM final metal concentration), 2.64 mol % biotin binding sites (corresponding to 916 μM biotin binding sites), and substrate **7** (34 mM final concentration, total reaction volume of 200 μL).

(Table 1, entries 1−4). These results are in agreement with the findings of Ogo and co-workers by which $[Cp^*Ir(H_2O)_3]^{2+}$ is a significantly more active transfer hydrogenation catalyst than $[Cp^*Rh(H_2O)_3]^{2+}$.[27]

(2) The presence of a histidine residue at position S112 Sav has a dramatic effect on the activity and the selectivity of the ATH. Salsolidine **8** is produced in 95% conversion and 41% ee (S) using $[\eta^5\text{-(Biot-2)}RhCl_2]$ **5** ⊂ S112H at 55 °C and pH 6.5. The activity and the selectivity can be further improved upon lowering the pH to 5.0 (quantitative conversion with 1 mol % catalyst **5** and 55% ee (S)) (Table 1, entries 5 and 6). In stark contrast, the presence of a histidine residue at position S112H has a detrimental effect on ATH with iridium (Table 1, entry 9). These results corroborate those of both Süss-Fink and Himeda whereby the transfer hydrogenase activity of $[Cp^*Rh(diimine)Cl]^+$ is reportedly higher than that of $[Cp^*Ir(diimine)Cl]^+$ (diimine = phenanthroline or 2,2′-bipyridine, respectively).[28,29]

(3) Introduction of other coordinating amino acids bearing soft donors (e.g., cysteine or methionine, Table 1, entries 15 and 18) at position S112 affords salsolidine in high yield albeit low enantioselectivity. Again here, we speculate that the metal is activated by coordination to a suitably positioned soft amino acid residue. Docking of $[\eta^5\text{-(Biot-2)}RhCl_2]$ **5** ⊂ S112 M indeed suggests that dative anchoring is likely (see SI).

(4) Introduction of a histidine residue at position K121H Sav affords (R)-salsolidine **8** in up to 79% ee using the rhodium catalyst at 55 °C (Table 1, entries 10 and 11). Again here, rac-**8** is produced with the iridium catalyst in the presence of K121H. Lowering the temperature leads, for both S112H and K121H mutants and in the presence

of the rhodium catalyst, to a decrease in enantioselectivity (Table 1, entries 7 and 8 and 12 and 13). This suggests that the $\Delta G^{\ddagger}$ term is dominated by the entropy contribution $\Delta S^{\ddagger}$.

(5) All point mutants tested devoid of a histidine at either position S112 or K121 afforded nearly racemic salsolidine **8** (ee <20%) (Table 1, entries 15−19). Interestingly, the double mutant S112H−K121H Sav gave very low levels of enantioselectivity (Table 1 entry 20). We speculate that in the presence of the double mutant, the biotinylated cofactor binds indiscriminately and, thus, in equimolar amounts to either of S112H and K121H histidines, resulting in the production of (rac)-**8**.

(6) Compared to $[\eta^5\text{-(Biot-2)}RhCl_2]$ **5** ⊂ WT Sav, both histidine bearing artificial metalloenzymes $[\eta^5\text{-(Biot-2)}RhCl_2]$ **5** ⊂ S112H Sav and $[\eta^5\text{-(Biot-2)}RhCl_2]$ **5** ⊂ K121H display increased turnover frequencies: 1 h^{-1} for the WT Sav compared to 6 h^{-1} for both S112H and K121H mutants.

Structural Characterization. Next, X-ray crystal structure analyses were carried out with the most selective artificial metalloenzymes: $[(\eta^5\text{-Biot-2})RhCl_2]$ ⊂ S112H and $[\eta^5\text{-(Biot-2)}RhCl_2]$ ⊂ K121H. Crystals of apo-Sav mutants S112H and K112H were soaked at pH 6 with an excess of complex $[\eta^5\text{-(Biot-2)}RhCl_2]_2$ **5$_2$**. The crystal structures of **5** ⊂ S112H and **5** ⊂ K121H could be solved to 2.4 and 1.8 Å resolution, respectively. Both structures contained strong electron density in the $F_o − F_c$ map in the biotin-binding pocket and in the proximity of the histidine residue either at position H112 or H121. The monomeric complexes $[\eta^5\text{-(Biot-2)}Rh]^{2+}$ and $[\eta^5\text{-(Biot-2)}RhCl]^+$ were fitted into the electron densities localized around H112 and H121′ respectively (Figure 2 and Figure S3).
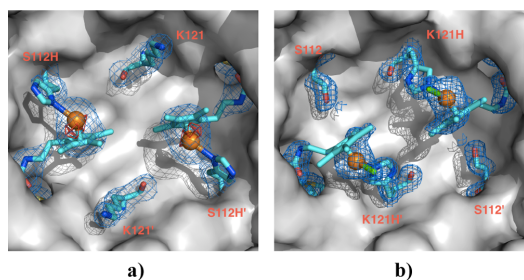
Figure 2. Close-up view of the X-ray crystal structure of $[(\eta^S\text{-Biot-}\mathbf{2})\text{Rh}]^{2+} \subset$ S112H Sav (a) and $[(\eta^S\text{-Biot-}\mathbf{2})\text{RhCl}]^+ \subset$ K121H Sav (b). The protein is displayed as surface representation and the ligand molecules as well as residues 112 and 121 in stick representation. The positions of ligand molecules are indicated by the $2F_o-F_c$ electron density (blue, contoured at 1.0 σ) and anomalous difference density (red, contoured at 4σ, due to weak absorption at a wavelength of 1.00 Å; no anomalous density was observed in structure $[(\eta^S\text{-Biot-}\mathbf{2})\text{RhCl}]^+$ $\subset$ K121H Sav). Rhodium is colored in orange and chlorine in green. Residues H121′ and K121′ are provided by the adjacent Sav monomer.

As highlighted by the docking simulations, the H121′ of an adjacent monomer in homotetrameric Sav lies closest to the Rh in the $\mathbf{5} \subset$ K121H structure, Figures 1c and 2b. Dative bonds were modeled between rhodium and the corresponding histidine: Rh−Nδ^{His112} 2.3 Å for $\mathbf{5} \subset$ S112H and Rh−N$\epsilon^{\text{His121′}}$ 2.3 Å for $\mathbf{5} \subset$ K121H. The Rh····Rh distances between two biotinylated molecules bound to neighboring symmetry-related Sav monomers are 9.5 and 9.2 Å for $\mathbf{5} \subset$ S112H and $\mathbf{5} \subset$ K121H, respectively. The position of the complexes in the S112H and K121H isoforms differs by an approximate 180° rotation of the Cp*Rh(Cl) group along an axis pointing toward the biotin binding pocket. Compared to the X-ray structure of biotin-loaded Sav (pdb code 1stp), the rmsd of the C$_\alpha$ for $\mathbf{5} \subset$ S112H and $\mathbf{5} \subset$ K121H is 0.247 and 0.214 Å, respectively, highlighting the minimal structural reorganization of the host protein upon introduction of the biotinylated cofactor.

For the complex bound to S112H, no additional electron density to complete the three-legged piano stool geometry could be detected, whereas the presence of a chloride ligand was apparent for the K121H structure.

Comparison of the docked vs X-ray structures of both mutant metalloenzymes reveals a semiquantitative agreement: the computed rmsd = 3.4 Å for the piano stool atoms in $\mathbf{5} \subset$ S112H and 2.1 Å for the piano stool atoms in $\mathbf{5} \subset$ K121H. Upon anchoring the catalyst via Rh−Nδ^{His112}, as found in the X-ray structure (Figure 2a), the rmsd decreases to 1.23 Å (see SI). These deviations can by-and-large be traced back to the differing conformation adopted by the ethylene spacer (docked vs X-ray, see SI). Despite this, the docking proved predictive enough to successfully engineer dative anchoring sites for the biotinylated cofactor. A full parametrization of the biotinylated Rh piano stool moiety will be required to yield more reliable docking results.

Inspection of the steric environment enforced by the host protein around the $[(\eta^S\text{-Biot-}\mathbf{2})\text{RhCl}_2] \subset$ S112H and $[\eta^S\text{-(Biot-}\mathbf{2})\text{RhCl}_2] \subset$ K121H reveals a "pseudomirror image relationship" between the two mutant artificial metalloenzymes, thus accounting for the production of (S)-$\mathbf{8}$ and (R)-$\mathbf{8}$, respectively (Figure 3a,b).
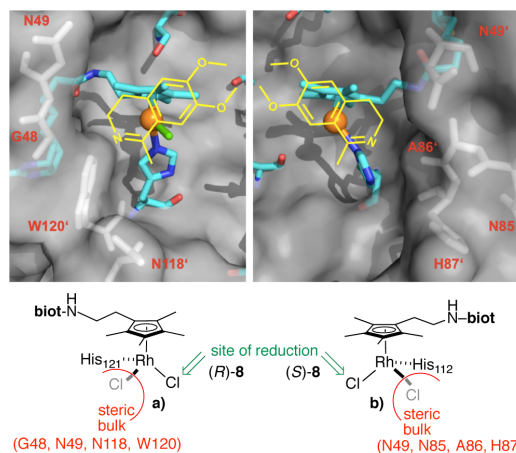
Figure 3. Schematic representation of $[(\eta^S\text{-Biot-}\mathbf{2})\text{RhCl}_2] \subset$ K121H Sav (a) and $[(\eta^S\text{-Biot-}\mathbf{2})\text{RhCl}_2] \subset$ S112H Sav (b). The steric bulk provided by Sav creates "pseudomirror-image environments" around the two transfer hydrogenases, thus favoring the approach of opposite prochiral faces of the imine substrate (in yellow).

OUTLOOK

Dual anchoring of a biotinylated three-legged rhodium piano stool complex affords an artificial transfer hydrogenase. It was shown that coordination of the catalyst precursor to a suitably positioned histidine residue has a significant impact on the catalyst's performance, both in terms of activity and of selectivity. The computational design toward the identification of the spacer length and site of mutation was confirmed by X-ray crystallography for the most selective artificial metalloenzymes $[\eta^S\text{-(Biot-}\mathbf{2})\text{RhCl}_2]$ $\mathbf{5} \subset$ K121H and $[\eta^S\text{-(Biot-}\mathbf{2})\text{RhCl}_2]$ $\mathbf{5} \subset$ S112H. Current efforts are aimed at computing the transition state to identify and engineer additional second coordination sphere interactions to improve the catalytic performance of these dually anchored artificial metalloenzymes.

ASSOCIATED CONTENT

Ⓢ Supporting Information

Modeling, biotinylated complex synthesis and characterization, catalytic runs, and X-ray data. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

thomas.ward@unibas.ch

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

REFERENCES

(1) Ward, T. R. *Acc. Chem. Res.* **2011**, *44*, 47.

5387

dx.doi.org/10.1021/ja309974s | *J. Am. Chem. Soc.* 2013, 135, 5384−5388

(2) Rosati, F.; Roelfes, G. *ChemCatChem* **2010**, *2*, 916.

(3) Deuss, P. J.; den Heeten, R.; Laan, W.; Kamer, P. C. *Chem.—Eur. J.* **2011**, *17*, 4680.

(4) Lu, Y.; Yeung, N.; Sieracki, N.; Marshall, N. M. *Nature* **2009**, *460*, 855.

(5) Abe, S.; Ueno, T.; Reddy, P. A. N.; Okazaki, S.; Hikage, T.; Suzuki, A.; Yamane, T.; Nakajima, H.; Watanabe, Y. *Inorg. Chem.* **2007**, *46*, 5137.

(6) Reetz, M. T. *Top. Organomet. Chem.* **2009**, *25*, 63.

(7) Wilson, M. E.; Whitesides, G. M. *J. Am. Chem. Soc.* **1978**, *100*, 306.

(8) Lin, C.-C.; Lin, C.-W.; Chan, A. S. C. *Tetrahedron: Asymmetry* **1999**, *10*, 1887.

(9) Collot, J.; Gradinaru, J.; Humbert, N.; Skander, M.; Zocchi, A.; Ward, T. R. *J. Am. Chem. Soc.* **2003**, *125*, 9030.

(10) Reetz, M. T.; Peyeralans, J. J.-P.; Maichele, A.; Fu, Y.; Maywald, M. *Chem. Commun.* **2006**, 4318.

(11) Hyster, T. K.; Knörr, L.; Ward, T. R.; Rovis, T. *Science* **2012**, *338*, 500.

(12) Klein, G.; Humbert, N.; Gradinaru, J.; Ivanova, A.; Gilardoni, F.; Rusbandi, U. E.; Ward, T. R. *Angew. Chem., Int. Ed.* **2005**, *44*, 7764.

(13) Dürrenberger, M.; Heinisch, T.; Wilson, Y. M.; Rossel, T.; Nogueira, E.; Knörr, L.; Mutschler, A.; Kersten, K.; Zimbron, M. J.; Pierron, J.; Schirmer, T.; Ward, T. R. *Angew. Chem., Int. Ed.* **2011**, *50*, 3026.

(14) (a) Carey, J. R.; Ma, S. K.; Pfister, T. D.; Garner, D. K.; Kim, H. K.; Abramite, J. A.; Wang, Z.; Guo, Z.; Lu, Y. *J. Am. Chem. Soc.* **2004**, *126*, 10812. (b) Garner, D. K.; Liang, L.; Barrios, D. A.; Zhang, J.-L.; Lu, Y. *ACS Catal.* **2011**, *1*, 1083.

(15) Ohashi, M.; Koshiyama, T.; Ueno, T.; Yanase, M.; Fujii, H.; Watanabe, Y. *Angew. Chem., Int. Ed.* **2003**, *42*, 1005.

(16) Creus, M.; Pordea, A.; Rossel, T.; Sardo, A.; Letondor, C.; Ivanova, A.; LeTrong, I.; Stenkamp, R. E.; Ward, T. R. *Angew. Chem., Int. Ed.* **2008**, *47*, 1400.

(17) Zimbron, J. M.; Sardo, A.; Heinisch, T.; Wohlschlager, T.; Gradinaru, J.; Massa, C.; Schirmer, T.; Creus, M.; Ward, T. R. *Chem.—Eur. J.* **2010**, *16*, 12883.

(18) Koehler, V.; Mao, J.; Heinisch, T.; Pordea, A.; Sardo, A.; Wilson, Y. M.; Knoerr, L.; Creus, M.; Prost, J.-C.; Schirmer, T.; Ward, T. R. *Angew. Chem., Int. Ed.* **2011**, *50*, 10863.

(19) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. *J. Comput. Chem.* **2005**, *26*, 1781.

(20) MacKerell, A. D., Jr.; Brooks, C. L., III; Nilsson, L.; Roux, B.; Won, Y.; Karplus, M. *CHARMM: The Energy Function and Its Parameterization with an Overview of the Program*; John Wiley & Sons: Chichester, 1998; Vol. 1.

(21) Schmid, M.; Nogueira, E. S.; Monnard, F. W.; Ward, T. R.; Meuwly, M. *Chem. Sci.* **2012**, *3*, 690.

(22) Izrailev, S.; Stepaniants, S.; Balsera, M.; Oono, Y.; Schulten, K. *Biophys. J.* **1997**, *72*, 1568.

(23) Van Leusen, D.; Beetstra, D. J.; Hessen, B.; Teuben, J. H. *Organometallics* **2000**, *19*, 4084.

(24) Reiner, T.; Jantke, D.; Raba, A.; Marziale, A. N.; Eppinger, J. *J. Organomet. Chem.* **2009**, *694*, 1934.

(25) Amouri, H. E.; Gruselle, M.; Jaouén, G. *Synth. React. Inorg. Met.-Org. Chem.* **2010**, *24*, 395.

(26) Kaufman, T. *Tetrahedron: Asymmetry* **2004**, *15*, 1203.

(27) Ogo, S.; Makihara, N.; Watanabe, Y. *Organometallics* **1999**, *18*, 5470.

(28) Canivet, J.; Süss-Fink, G.; štěpnička, P. *Eur. J. Inorg. Chem.* **2007**, *2007*, 4736.

(29) Himeda, Y.; Onozawa-Komatsuzaki, N.; Sugihara, H.; Arakawa, H.; Kasuga, K. *J. Mol. Catal. A: Chem.* **2003**, *195*, 95.

87

Supporting Information for

A dual anchoring strategy for the localization and activation of artificial metalloenzymes based on the biotin-streptavidin technology

Jeremy M. Zimbron, Tillmann Heinisch, Maurus Schmid, Didier Hamels, Elisa S. Nogueira, Tilman Schirmer and Thomas R. Ward*

Biozentrum and Department of Chemistry, University of Basel CH-4056 Basel, Switzerland

e-mail: thomas.ward@unibas.ch

Docking

The structure of the monomeric biotinylated complexes $[\eta^5$-(Biot-**2**)$RhCl_2(H_2O)]$ and $[(\eta^5$-Biot-L)$RhCl_2]$ **5** were optimized with Density Functional Theory (DFT[1]) using Gaussian[2] and the hybrid B3LYP functional.[3,4] The atomic basis sets consisted of an effective core potential (LanL2DZ[5]) for the metal and the explicit 6-31G(d,p) basis for all remaining atoms.

Mutations were introduced using VMD[6] at either positions S112H or K121H respectively. Simulations were performed in NAMD[7] using the CHARMM2[7] force-field[8] with custom parameters developed in our group[9] or elsewhere[10]. No dihedral parameters involving the rhodium center were considered. Equilibrium distances around the metal center were set to the DFT values with an estimated force constant. As the geometry around the metal is conserved, these parameters are sufficient to maintain the pianostool arrangement. Charges were taken from the DFT calculation above (as Mulliken charges). The Lennard-Jones parameters for the metal were taken from the UFF force-field.[11]

The system was set up using VMD11. An additional bond from the metal to the histidine nitrogen N_ε was introduced for the simulation of the mutated proteins to mimic the histidine interaction with the metal.

S2
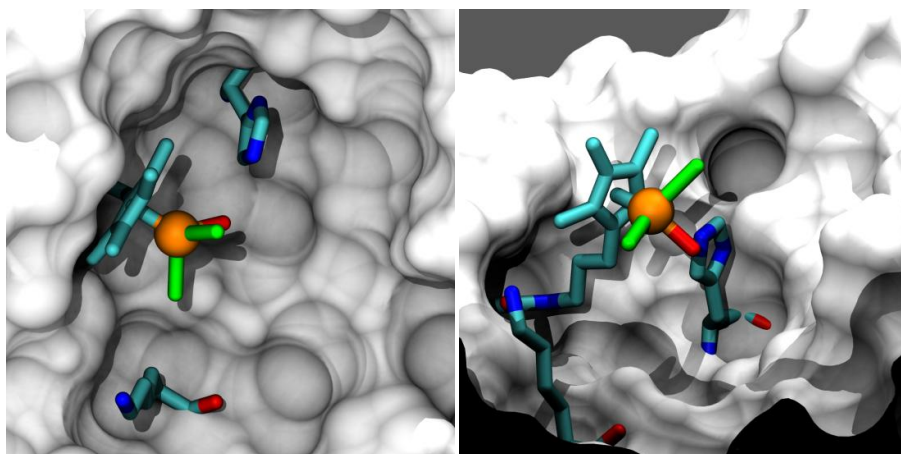
a)

b)
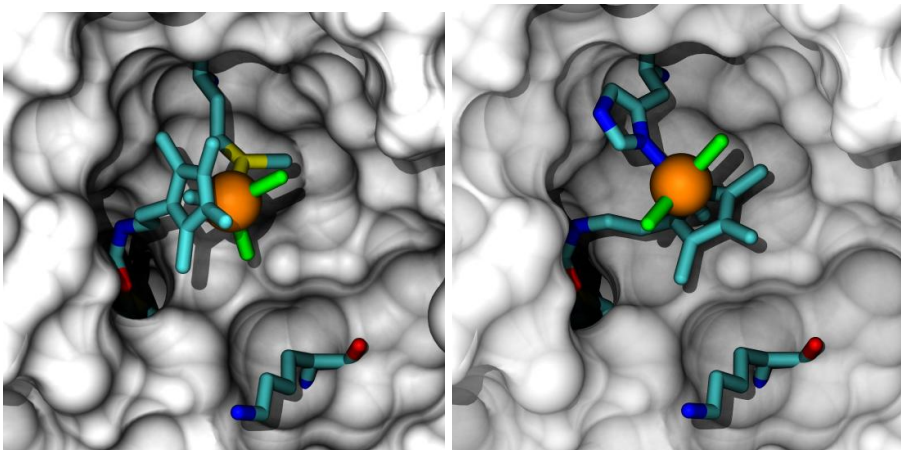
c)

d)
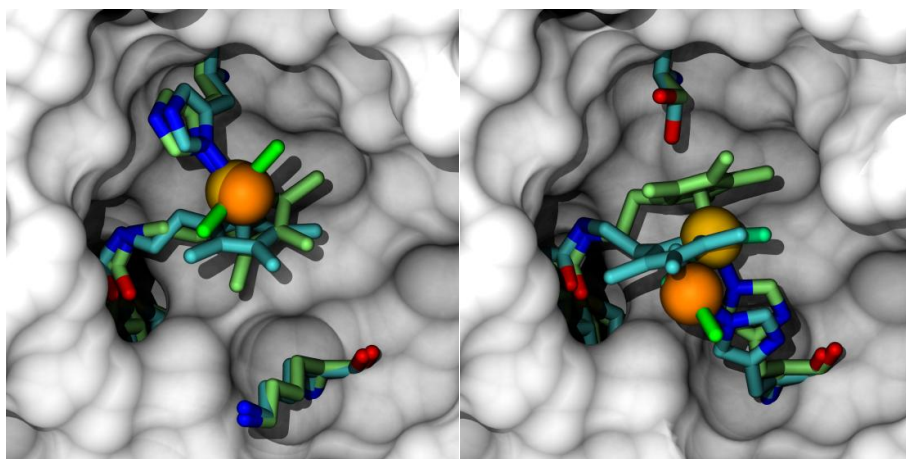
e)

f)

SI Figure 1. Docked structure of [η⁵-(Biot-CH₂Cp*)RhCl₂(H₂O)] ⊂ S112H Sav a); [(η⁵-Biot(CH₂)₃Cp*)RhCl₂] **5** ⊂ S112H Sav b); [η⁵-(Biot-2)RhCl₂] **5** ⊂ S112M c); [η⁵-(Biot-2)RhCl₂] **5** ⊂ S112H(ᴺδ) d); overlay of docked (lime) *vs.* X-ray (cyan) of [η⁵-(Biot-2)RhCl₂] **5** ⊂ S112H and e) [η⁵-(Biot-2)RhCl₂] **5** ⊂ K121H f) The pianostool complex and streptavidin's side-chains at positions 112 and 121 are displayed as stick representation, Sav as solvent accessible surface.


Synthesis

2-(2,3,4,5-tetramethylcyclopentadienyl)ethylamine (**2**) was synthesized as previously described.[12]

[η⁵-(**2**)RhCl₂]₂²⁺ **3** was synthesized as previously described.[13]

[η⁵-(**2**)IrCl₂]₂²⁺ **4** was synthesized according to Jaouen *et al.* protocol.[14]
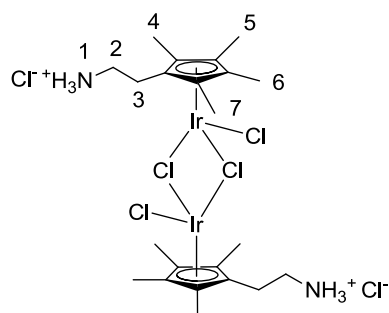
2-(2,3,4,5-tetramethylcyclopentadienyl)ethyl amine **2** (tautomers mixture, 91.1 mg, 0.551 mmol) was dissolved in Et₂O (2 mL) and treated with HCl (0.272 mL, 1.09 mmol, 4 M in dioxane). Hexane (4 mL) was added to this solution until a white precipitate formed, which was filtered and washed twice with hexane. The white solid was dissolved in MeOH (5 mL) and introduced into a 10 mL Schlenk tube. Then, [Ir(COD)Cl]₂ (100 mg, 0.149 mmol) and a solution of concentrated HCl (0.8 mL of 37% solution in H₂O) was added. This mixture was refluxed under nitrogen; after 30 min the initial orange-red suspension turned to a clear orange solution. A few minutes later, orange crystals began to precipitate and the reaction mixture was refluxed overnight. The precipitate was filtered off, washed with MeOH and dried to yield a bright orange powder (116.8 mg, 60% yield).

¹H NMR (400 MHz, DMSO-d_6) δ 8.23 (br s, 6H, **H¹**), 2.95-2.84 (m, 4H, **H²**), 2.43-2.39 (m, 4H, **H³**), 1.74/1.64 (s, 24H, **H⁴**, **H⁵**, **H⁶**, **H⁷**). ¹³C NMR (101 MHz, DMSO-d_6) δ 97.34, 92.50, 86.22, 37.10, 22.32, 9.23 (2C), 9.11 (2C). HR-MS (ESI-MS, pos) calcd. for C₂₂H₃₆Cl₄N₂Rh₂, 854.0891. *m/z* = 819.1130 [M-Cl]⁺, 642.22 [(M-Cl)/2]⁺.

[η⁵-(Biot-**2**)RhCl₂]₂ (**5₂**)
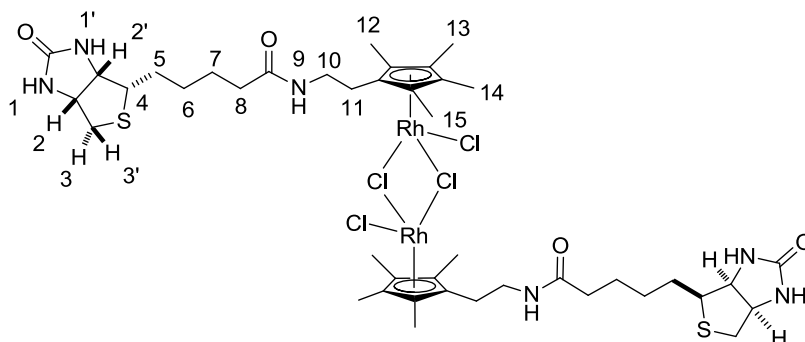
[η⁵-(**2**)RhCl₂]₂²⁺ **3** (200 mg, 0.296 mmol) was suspended in DMF (20 mL) and biotin pentafluorophenylester (231 mg, 0.562 mmol) and Et₃N (0.831 mL, 5.920 mmol) were added. The compound rapidly dissolved to give a dark red solution and was stirred at RT overnight. The solvent was removed and the reddish solid was suspended in hexane. The solid was filtered and washed with hexane (3x15 mL) and then with CH₂Cl₂ (4x15 mL) to afford a brown-red powder (284 mg, 85% yield).

¹H NMR (500 MHz, DMSO-d_6) δ 7.97 (t, *J* = 6.0 Hz, 2H, **H⁹**), 6.44 (s, *2*H, **H¹′**), 6.37 (s, *2*H, **H¹**), 4.35-4.27 (m, 2H, **H²**), 4.15-4.09 (m, 2H, **H²′**), 3.22 (dt, *J* = 13.4, 6.9 Hz, 4H, **H¹⁰**), 3.11-3.09 (m, 2H, **H⁴**), 2.83 (dd, *J* = 12.4, 5.0 Hz, 2H, **H³′**), 2.59 (d, *J* = 12.4 Hz, 2H, **H³**), 2.27 (t, *J* = 6.9 Hz, 4H, **H¹¹**), 2.03 (t, *J* = 7.4 Hz, 4H, **H⁸**), 1.68/1.64 (s, 24H, **H¹²**, **H¹³**, **H¹⁴**, **H¹⁵**), 1.80-1.21 (m, 12H, **H⁵**, **H⁶**, **H⁷**). ¹³C NMR (126 MHz, DMSO-d_6) δ 172.05 (2C), 162.68 (2C), 100.40 (4C), 99.04 (4C), 96.78 (2C), 61.01 (2C), 59.15 (2C), 55.37 (2C), 54.89 (2C), 35.65 (2C), 35.05 (2C), 28.19 (2C), 28.00 (2C), 25.07 (2C), 24.13 (2C), 8.70 (4C), 8.59 (4C). HR-MS (ESI-MS, pos) calcd. for C₄₂H₆₄Cl₄N₆O₄Rh₂S₂, 1126.1451: *m/z* = 1093.1555 [M-Cl]⁺, 492.12 [(M-2Cl)/2]²⁺, 528.10 [(M-Cl)/2]⁺.

[η⁵-(Biot-**2**)IrCl₂]₂ **6₂**
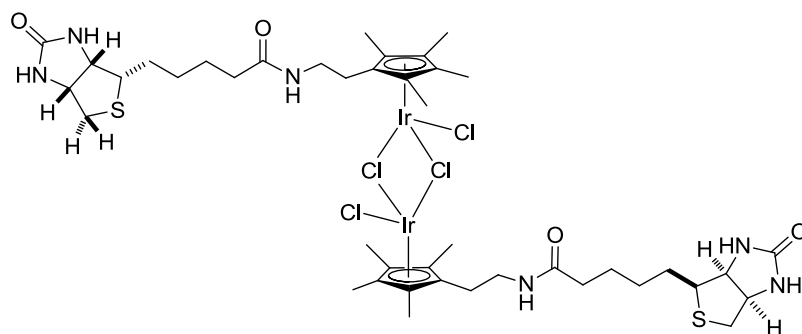
$[\eta^5\text{-}(\mathbf{2})\text{IrCl}_2]_2{}^{2+}$ **4** (200 mg, 0.234 mmol) was suspended in DMF (20 mL) and biotin pentafluorophenyl ester (182 mg, 0.445 mmol) and Et_3N (0.658 μL, 4.68 mmol) was added. The compounds rapidly dissolved to give a clear yellow solution. The reaction was stirred at RT overnight. The solvent was removed and the resulting brown solid was suspended in hexane. The precipitate was filtered and washed with hexane (3x15 mL) then with CH_2Cl_2 (4x15 mL) to afford a pale yellow powder (66 mg, 80% yield).

^{1}H NMR (400 MHz, DMSO-d_6) δ 7.94 (t, J = 5.9 Hz, 2H, $\mathbf{H^9}$), 6.41 (s, *2H*, $\mathbf{H^{1'}}$), 6.35 (s, *2H*, $\mathbf{H^1}$), 4.32-4.24 (m, 2H, $\mathbf{H^2}$), 4.11-4.09 (m, 2H, $\mathbf{H^{2'}}$), 3.19 (dt, J = 13.2, 6.9 Hz, 4H, $\mathbf{H^{10}}$), 3.14-3.02 (m, 2H, $\mathbf{H^4}$), 2.80 (dd, J = 12.5, 5.0 Hz, 2H, $\mathbf{H^{3'}}$), 2.55 (d, J = 12.5 Hz, 2H, $\mathbf{H^3}$), 2.15 (t, J = 6.9 Hz, 4H, $\mathbf{H^{11}}$), 2.02 (t, J = 7.4 Hz, 4H, $\mathbf{H^8}$), 1.67/1.61 (s, 24H, $\mathbf{H^{12}}$, $\mathbf{H^{13}}$, $\mathbf{H^{14}}$, $\mathbf{H^{15}}$), 1.73-1.20 (m, 12H, $\mathbf{H^5}$, $\mathbf{H^6}$, $\mathbf{H^7}$). ^{13}C NMR (126 MHz, DMSO-d_6) δ 171.99 (2C), 162.76 (2C), 94.45 (4C), 92.12 (4C), 88.99 (2C), 61.01 (2C), 59.15 (2C), 55.59 (2C), 39.61 (2C), 35.91 (2C), 35.14 (2C), 28.15 (2C), 27.93 (2C), 25.25 (2C), 23.95 (2C), 8.36 (4C), 8.24 (4C). HR-MS (ESI-MS, pos) calcd. for $C_{42}H_{64}Cl_4Ir_2N_6O_4S_2$, 1306.2443. m/z = 1271.2703 [M-Cl]$^+$, 618.1216 [M/2-Cl]$^+$.

Streptavidin Production

Mature streptavidin[15] and mutants thereof were expressed and purified according to our published protocol.[16] All proteins were characterized by ESI-MS and their biotin binding affinity was determined using the biotin-4-fluorescein assay developed by Gruber.[17]

Catalysis Experiments

All experiments were carried out using standard Schlenk techniques, with thoroughly degassed solutions (nitrogen-flushed).

Preparation of Stock Solutions

Sodium formate (3.4 g, 0.05 mol, 4 M) and MOPS (8.37 g, 0.04 mol, 3.2 M) were dissolved in milli-Q water (12.5 mL) and the pH was adjusted to 5.0 using concentrated HCOOH solution. The dimeric metal complexes **5**$_2$ and **6**$_2$ were dissolved in DMSO (0.05 M) and stored at 4°C prior to use. A stock solution of the substrate was prepared dissolving 6,7-dimethoxy-1-methyl-3,4-dihydroisoquinoline **7** in milli-Q water to a final concentration of 2 M.

Procedure for the catalytic reduction of imine 7 using 5 ⊂ S112H and 5 ⊂ K121H at pH 5.0 with 1 mol% Metal Loading

The streptavidin isoform was directly weighed into the reaction tubes (0.916 mM final biotin binding site concentration). A solution of formate/MOPS (193 μL) from the stock solution was added. The mixture was stirred for 5-10 min until the protein was completely dissolved. The dimer precursor complex from the stock solution was added: 0.92 μL, final metal concentration, 458 μM; 0.5 equivalents [Rh] *vs*. Sav free binding sites. The reaction mixture was stirred for 15 min. Finally, the substrate stock solution was added (4.58 μL, final concentration, 45.8 mM). The test tubes were placed in a magnetically stirred multireactor (RR 98072, Radleys Discovery Technologies) which was purged three times with nitrogen heated to the appropriate temperature and time.

For other catalysis experiments refer to the catalytic conditions in the main text (see Table **1)**
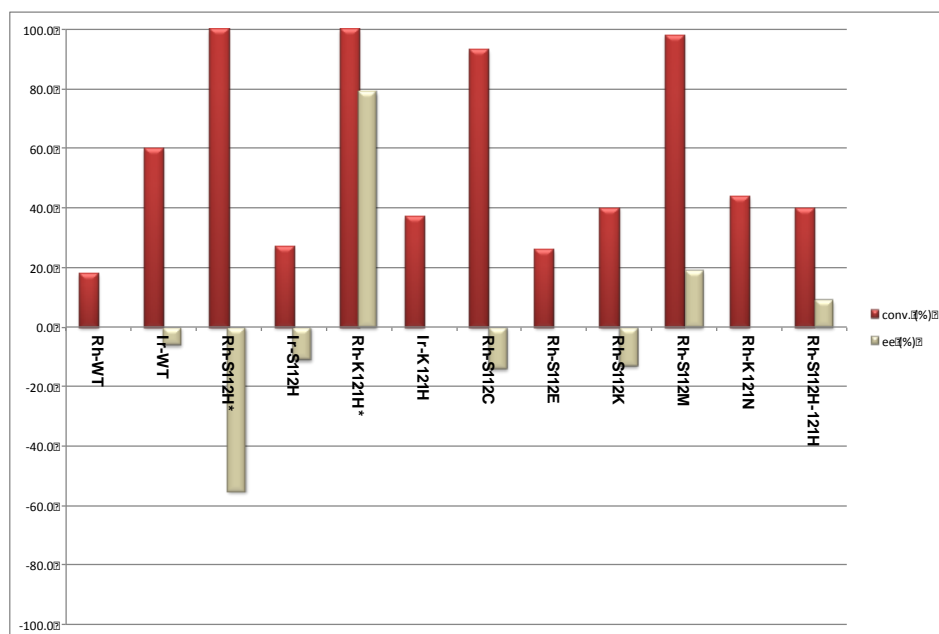
SI Figure 2. Graphical summary of catalytic results: conversion (red bars), ee (beige bars), * represent the results under the optimized conditions , See Table 1 for details.

X-ray Crystal Structures

Protein Crystallization and Biotinylated Complex Soaking. Crystals of streptavidin mutants Sav K121H and Sav S112H (residues 13-159 of mature streptavidin fused N-terminally to the 12 residues of a T7-tag)[18] were obtained within one month by hanging-drop vapor diffusion technique mixing 2 μL crystallization buffer (0.1 M MES, 19 % PEG500, pH 6.0) and 3 μL protein solution (20 mg/mL in water) prior to equilibration of the solution against a reservoir of 500 μL crystallization buffer. Subsequently, single crystals of mutants Sav K121H and Sav S112H were soaked for 4 h at room temperature in a soaking buffer, which was prepared by mixing 1 μL of a 10 mM stock solution of complex **5**$_2$ ([η^5-(Biot-**2**)RhCl$_2$]$_2$) in dimethylformamide and 9 μL crystallization buffer. After the soaking, crystals were shock-frozen in liquid nitrogen.

Data Processing and Refinement. X-ray diffraction data

were collected at the Swiss Light Source beam line X06DA and processed with software XDS[19] or MOSFLM (CCP4 Suite)[20]. The structures were solved by molecular replacement using program PHASER[21] (CCP4 Suite) and the structure 2qcb from the PDB as model input with ligand and water molecules removed. For structure refinement PHENIX.REFINE[22] was used. Crystallographic details are given in Supporting Table 1. Structural figures were drawn with software DINO. The structures were deposited in the pdb repository with the codes 4GJS and 4GJV for $[\eta^5$-(Biot-**2**)RhCl]$^+$ $\subset$ K121H and $[\eta^5$-(Biot-**2**)Rh]$^{2+}$ $\subset$ S112H respectively. Supporting Table 1: Summary of data processing and refinement.

Data processing		
Complex	$[\eta^5$-(Biot-**2**)RhCl]$^+$ $\subset$K121H	$[\eta^5$-(Biot-**2**)Rh]$^{2+}$ $\subset$ S112H
Wavelength of X-ray experiment (Å)	1.00	1.55
Software	MOSFLM	XDS
Resolution (Å)	40.80 - 1.85	150.00 - 2.40
Space group	C2	I4₁22
Cell dimensions (Å)	a=81.7, b=81.3, c=47.0, β=104.8	a=57.6, b=57.6, c=183.8
R_{merge}	8.1 (18.5)	9.8 (79.2)
No unique reflections	24972	10085
Multiplicity	2.7	3.4
I/sig(I)	6.0 (2.3)	10.3 (1.6)
Completeness (%)	98.5 (96.3)	86.2 (73.2)
CC (1/2)	n/d	62.5
Refinement		
Resolution (Å)	28.32 - 1.85	19.88 - 2.40
R_{work}/R_{free} (%)	20.3 (25.0)	18.5 (23.0)
$RMSD_{bond}$ (Å)	0.007	0.032
$RMSD_{angle}$ (deg)	1.365	1.121
Bfactor all atoms solvent $[\eta^5$-(Biot-**2**)RhCl]$^+$ /$[\eta^5$-(Biot-**2**)Rh]$^{2+}$	23 33 32	36 33 44

Values in brackets refer to the highest resolution shell

Structure of $[\eta^5$-(Biot-2)RhCl]$^+$ $\subset$ Sav K121H

There are two streptavidin monomers in the asymmetric unit from which a tetramer can be formed by application of the crystallographic 2-fold symmetry axis parallel b (**Supporting Figure 1a**). The 12 N-terminal residues of the T7-tag and 25 residues at the C-terminus are not resolved, probably due to disorder. Strong residual electron density in the Fo-Fc omit map was present in the biotin binding pocket and its vestibule reaching to residue His121' of the neighboring SAV monomer suggesting the cleavage of Rh-dimer **5$_2$** upon binding to Sav K121H. The initial ligand model to be fit in the residual electron density comprised a rhodium-bound pentamethylcyclpentadienyl ring (Cp*Rh group) which was covalently attached to biotin via an ethylene spacer and an amide bond. The geometry of the [η^5-(Biot-**2**)Rh]$^{2+}$ moiety was derived from the X-ray structure of the closely related small molecule rhodium complex 1-(4-nitrophenyl)-1H-imidazol-3-yl)-dichloro-(η^5-pentamethylcyclopentadienyl)-rhodium (CSD code FACHEY, **Supporting Figure 1e, right side**). Ligand restraints were adapted manually using software REEL (PHENIX package). The biotin moiety is bound to the streptavidin pocket as described by Weber *et al.*[23] An H-bond is formed between the ligand's carbonyl group and the backbone nitrogen of residue N49. Due to structural heterogeneity, the ethylene spacer and portions of the aromatic 5-membered ring of complex [η^5-(Biot-**2**)Rh]$^{2+}$ are not properly resolved in the electron density (**Supporting Figure 1b**). The position of the rhodium atom as bound to the Cp* ring is well defined by the strong residual positive electron density signal (12 σ) in bonding distance (2.3 Å) to the Nε of His121' of the neighboring SAV monomer. Close to the metal, residual density suggested the presence of an additional chloride ligand. There is no density at the third rhodium coordination site. In the refinement, bond angles Cp*^{centroid}-Rh-Cl, Cp*^{centroid}-Rh-Nε$^{H121'}$ and Cl-Rh-Nε$^{H121'}$ were restrained to the values given by the X-ray structure of the related small molecule pianostool complex with CSD code FACHEY (**Supporting Figure 1e**).

Strong density (11 σ) was found in the Fo-Fc map close to His127 and was modeled as a rhodium ion covalently attached to the histidine-Nε, suggesting the binding of an additional rhodium, due to the use of a large excess biotinylated complex **5$_2$** for soaking. Finally, the

96

structure contains 328 water molecules.

Structure of $[\eta^5\text{-(Biot-2)Rh}]^{2+} \subset$ Sav S112H

There is one streptavidin molecule in the asymmetric unit. The SAV tetramer is obtained by application of the crystallographic dyads parallel to c and along the a-b diagonal. As in the structure of complex $[\eta^5\text{-(Biot-}\mathbf{2)}\text{RhCl}]^+ \subset$ Sav K121H, 12 N-terminal and 25 C-terminal amino acids are not resolved. Strong residual electron density in the Fo-Fc omit map was evident in the biotin-binding pocket and its vestibule close to the imidazole ring of His112 of the same subunit (**Supporting Figure 2c**). As expected from the small f'' (3.6 e-) of Rh at the employed wavelength, only a weak peak (5.3 σ) in the anomalous difference map was found in the vestibule in the same position close to residue H112. Complex $[\eta^5\text{-(Biot-}\mathbf{2)}\text{Rh}]^{2+}$ was modeled in the densities (**Supporting Figure 2d**). As in the structure of $[\eta^5\text{-(Biot-}\mathbf{2)}\text{RhCl}]^+ \subset$ Sav K121H, the ethylene spacer and the methyl-substituted 5-membered ring of complex $[\eta^5\text{-(Biot-}\mathbf{2)}\text{Rh}]^{2+}$ are not resolved in the electron density. The distance between the Cp*-bound Rh and $N\delta^{H112}$ is 2.3 Å indicating a coordinative bond. No further electron density was detected in the coordination sphere of the rhodium. However, the positions are likely to be occupied by two water molecules, which would give rise to a pianostool-type of ligand coordination sphere around the metal.

Additional strong electron density in the Fo-Fc omit map was found next to the following protein nitrogen atoms: backbone G48(N) (5 σ), K80(Nζ) (3 σ), R84(Nη) (7 σ), H87(Nε) (7 σ) and H127(Nε) (14 σ, 4.6 σ peak in anomalous scattering density). A rhodium atom was modeled next to His127 and chloride in all other positions. Again here this is most probably due to the use of a large excess biotinylated complex $\mathbf{5}_2$ for soaking purposes. Besides these ligands, the structure contains 23 water molecules.

a

b

c

d

Ligand **9** (red lines)

CSD code FACHEY

e
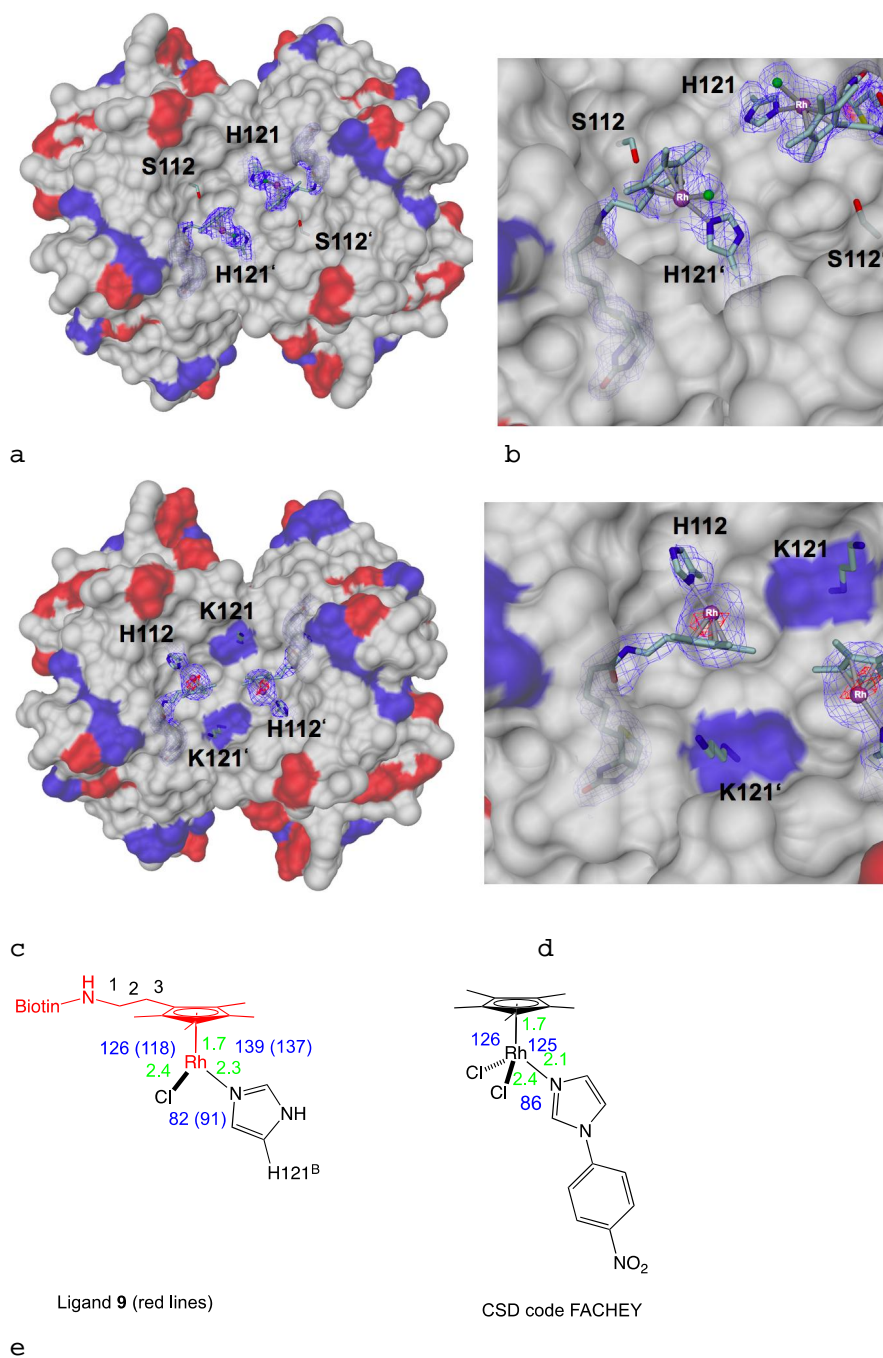
SI Figure 3: X-ray structures of $[\eta^5\text{-(Biot-}\mathbf{2})\text{RhCl]}^+ \subset$ K121H and $[\eta^5\text{-(Biot-}\mathbf{2})\text{Rh]}^{2+} \subset$ S112H. a) Top view of two cis-related biotin-binding pockets of tetrameric complex $[\eta^5\text{-(Biot-}\mathbf{2})\text{RhCl]}^+ \subset$ K121H in surface representation

(red: acidic amino acids, blue: basic amino acids). The 2Fo-Fc map covering complex [η^5-(Biot-**2**)RhCl]$^+$ and His121 is contoured at 1.2 σ. b) Close-up view of the ligand binding site of complex [η^5-(Biot-**2**)RhCl]$^+$ ⊂ K121H. c) [η^5-(Biot-**2**)Rh]$^{2+}$ ⊂ S112H, view as in a). d) Close-up view of complex [η^5-(Biot-**2**)Rh]$^{2+}$ ⊂ S112H. Significant anomalous scattering density (contoured at 4 σ, red) indicates the position of the rhodium. e) Comparison of the piano stool structures of complex [η^5-(Biot-**2**)RhCl]$^+$ as bound to Sav K121H (left) and CSD ligand FACHEY (right). Green and blue colored numbers indicate bond distances and angles, respectively (numbers in parentheses refer to symmetry-related ligand [η^5-(Biot-**2**)RhCl]$^+$ in same asymmetric unit).

SI Literature Cited

(1) Hohenberg, P.; Kohn, W. *Phys. Rev. B* **1964**, *136*, B864.

(2) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. In *Gaussian 09, Revision B.01, Gaussian, Inc., Wallingford CT* Wallingford CT, 2009.

(3) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. *J. Phys. Chem.* **1994**, *98*, 11623.

(4) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648.

(5) Hay, P. J.; Wadt, W. R. *J. Chem. Phys*. **1985**, *82*, 270.

(6) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics Modell*. **1996**, *14*, 33.

(7) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. *J. Comput. Chem*. **2005**, *26*, 1781.

(8) MacKerell Jr., A. D.; Brooks III, C. L.; Nilsson, L.; Roux, B.; Won, Y.; Karplus, M. *CHARMM: The Energy Function and Its Parameterization with an Overview of the Program*; John Wiley & Sons: Chichester, 1998; Vol. 1.

(9) Schmid, M.; Nogueira, E. S.; Monnard, F. W.; Ward, T. R.; Meuwly, M. *Chem. Sci*. **2012**, *3*, 690.

(10) Izrailev, S.; Stepaniants, S.; Balsera, M.; Oono, Y.; Schulten, K. *Biophys. J*. **1997**, *72*, 1568.

(11) Rappe, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard, W. A.; Skiff, W. M. *J. Am. Chem. Soc*. **1992**, *114*, 10024.

(12) Van Leusen, D.; Beetstra, D. J.; Hessen, B.; Teuben, J. H. *Organometallics* **2000**, *19*, 4084.

(13) Reiner, T.; Jantke, D.; Raba, A.; Marziale, A. N.; Eppinger, J. *J. Organomet. Chem*. **2009**, *694*, 1934.

(14) Amouri, H. E.; Gruselle, M.; Jaouén, G. *Synth. React. Inorg. Met.-Org. Chem*. **2010**, *24*, 395.

(15) Sano, T.; Cantor, C. R. *Proc. Natl. Acad. Sci. USA* **1990**, *87*, 142.

(16) Humbert, N.; Zocchi, A.; Ward, T. R. *Electrophoresis* **2005**, *26*, 47.

(17) Gruber, H. J.; Kada, G.; Marek, M.; Kaiser, K. *Biochim. Biophys. Acta* **1998**, *1381*, 203.

(18) Humbert, N.; Schurmann, P.; Zocchi, A.; Neuhaus, J. M.; Ward, T. R. *Methods Mol Biol* **2008**, *418*, 101.

(19) Kabsch, W. *Acta Crystallogr D* **2010**, *66*, 125.

(20) Collaborative *Acta Crystallogr., Sect. D: Biol. Crystallogr*. **1994**, *50*, 760.

(21) McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J. *J. Appl. Cryst* **2007**, *40*, 658.

(22) Adams, P. D.; Afonine, P. V.; Bunkóczi, G.; Chen, V. B.; Davis, I. W.; Echols, N.; Headd, J. J.; Hung, L.-W.; Kapral, G. J.; Grosse-Kunstleve, R. W.; McCoy, A. J.; Moriarty, N. W.; Oeffner, R.; Read, R. J.; Richardson,

D. C.; Richardson, J. S.; Terwilliger, T. C.; Zwart, P. H. *Acta Crystallogr. D Biol. Crystallogr.* **2010**, *66*, 213.

(23) Weber, P. C.; Ohlendorf, D. H.; Wendoloski, J. J.; Salemme, F. R. *Science* **1989**, *243*, 85

6. Simulating a Transition Metal Moiety

Chapter 5 demonstrated the usefulness of calculations for a rational understanding of the binding mode of metal complexes in artificial metalloenzymes. However, a more precise force field is required for more reliable simulations of whole artificial metalloenzymes or metal complex inhibited enzymes.

The Cambridge Structural Database (CSD[161]) is a repository for structural information of small molecules and metal complexes, similar to what the PDB is for proteins. Until now (beginning of 2013), it contains information about almost 550'000 structures, including more than 300'000 structures containing a transition metal.[a] The information available within is valuable for comparing calculated structure to experimental data. We selected a representative number of d^6-piano stool complexes to parametrize the force fields and compare the optimzed structures to these references.

As detailed in the introduction, force field simulation of metal complexes is not straight-forward. The implementation of Valbond into the CHARMM force field (Valbond-trans CHARMM) was available and had been employed previously,[95,162] however modifications to the approach were necessary to be able to incorporate d^6-piano stool complexes into the simulation. As outlined in the introduction, the η^n moiety present in piano stool complexes makes this a difficult task. Several approaches for simulating the η^n with $n = 5, 6$ moiety had been tested: (a) an all-atom-bonded approach with bonds from the metal to all aromatic carbon atoms concerned was evaluated. But the large number of bonds on the metal (9 for η^6 and 8 for η^5 piano stools) yielded problems with the

---

[a]http://www.ccdc.cam.ac.uk/products/csd/statistics/

force field: because there are different equilibrium angles between a piano stool leg and the equivalent carbons, several equilibrium angles for the same atom types needed to be considered. Another problem was the now four-bonding ring atoms causing problems for the planar sp^2 description required in Valbond. (b) a partly bonded (e.g. three bonds for the η^6) was not pursued anymore because of its implication on symmetry, as the six carbon atoms would not have been equivalent anymore (additionally to the same angle problem as above). (c) finally, an approach involving a dummy atom in the center of the arene moiety was implemented. Meanwhile, bugs in the Valbond code were fixed and the module was included into the most recent version of CHARMM (CHARMM37) to be available for public in the next release.

The findings have been pulished in the Journal of Chemical Theory and Computation (JCTC) and are attached below.[163]

6.1. A Broadly Applicable Force Field for d^6-Piano Stool Complexes

Toward a Broadly Applicable Force Field for d⁶-Piano Stool Complexes

Maurus H. Schmid,[†,‡] Thomas R. Ward,[*,†] and Markus Meuwly[*,‡]

[†]Department of Chemistry, University of Basel, Spitalstrasse 51, CH-4056 Basel, Switzerland

[‡]Department of Chemistry, University of Basel, Klingelbergstrasse 80, CH-4056 Basel, Switzerland

S *Supporting Information*

ABSTRACT: Three-legged piano stool complexes are prototypical organometallic complexes relevant to a wide range of chemically relevant questions. Force field parametrization of transition-metal complexes is difficult and underdeveloped, and metal-specific force fields and software are required. Here we report our efforts to derive parameters for the conventional CHARMM and the Valbond-CHARMM force fields for d⁶-piano stool complexes. In Valbond-CHARMM, the usual angular term is replaced with hybrid orbital strength functions. These functions describe the energy not only of distorted bond angles around the minimum but also at very large distortions. Structure optimizations led to a good agreement between the calculated force field and the X-ray structures. They were comparable to RMSDs obtained between X-ray and DFT structures. In addition, and contrary to treating the systems with DFT, molecular dynamics simulations on the multiple nanosecond time scale are possible and allow to compute meaningful structural and energetic observables. Explicit solvent simulations of the complexes in methanol and water allow to determine the solvent distribution around the complexes. The parametrization presented here will be a useful starting point for dynamics investigations of catalysts in structurally more demanding environments.

INTRODUCTION

Three-legged piano stool complexes are prototypical organometallic complexes. Over the years they have received increasing attention as catalyst precursors as well as organometallic building blocks which can act either as enzyme inhibitors or metalloenzyme mimetics.[1−4] To complement experiments, interpret observations, or even predict experimental findings, computational modeling is a valuable and often indispensable tool. Computational investigations of these complexes have traditionally used ab initio methods or density functional theory (DFT). However, with in silico high-throughput screening approaches in mind, the computational power required even with DFT becomes prohibitive. For studies of organometallic complexes together with metalloenzymes, mixed quantum mechanical/molecular mechanics (QM/MM) procedures may be attractive, but for library-screening purposes such an approach is not sufficiently efficient and scalable. Instead, force field methods would be highly desirable. As force field calculations are much faster than any of the QM methods, they not only allow screening of large libraries but also explicit atomistic simulations of much larger systems involving proteins including artificial metalloenzymes or enzymes with transition-metal inhibitors mentioned above.

Most general purpose force fields for atomistic simulations (e.g., CHARMM,[5] Amber,[6] Gromos[7] or OPLS[8]) are defined as a sum of bonded and nonbonded energies:

$$V = E_{bonded} + E_{nonbonded} \tag{1}$$

The bonded energies are sums over harmonic potentials for stretching and bending terms and periodic functions for torsions:

$$
\begin{aligned}
E_{bonded} =& \sum_{bonds} k_r(r - r_0)^2 + \sum_{angles} k_\theta(\theta - \theta_0)^2 \\
&+ \sum_{torsion} k_\gamma(1 + \cos(n\omega - \gamma)) \\
&+ \sum_{improper} k_\phi(\phi - \phi_0)^2
\end{aligned}
\tag{2}
$$

Additional terms such as Urey−Bradley terms are commonly included as well. The nonbonded terms include sums over electrostatic (point charge) and vdW terms.

$$E_{nonbonded} = \sum_{i<j} \left\{ 4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^6 \right] + \frac{q_i q_j e^2}{r_{ij}} \right\} \tag{3}$$

Metallocene and bent-metallocene complexes have been parametrized in a few instances.[9−11] But generally, force field parametrization of transition-metal complexes is difficult and underdeveloped, and metal-specific force fields and software are required,[12−19] as it has been suggested that the harmonic

approximation for bonds and angles may not be the most suitable for transition metals.[17,18,20,21]

In order to describe ground-state geometries, a considerable number of force field parameters is required: every bond length, angle, and dihedral angle requires at least two parameters. In this context, the Valbond (VB)[22] formalism alleviates some of these drawbacks for the angular terms. Instead of two parameters (force constant and equilibrium angle) for each atom type triad, VB requires a hybridization for each atom, which is usually determined automatically but can be assigned by the user, and only one parameter per element pair to describe the bond hybridization. This reduces the parametrization effort, especially for complicated structures with many different atom types. For example, for the most simple complex **1** (see Figure 1), 6 VB instead of 20 CHARMM angle parameters are needed. For the more complex case of compound **4**, VB requires 7 parameters, whereas CHARMM requires more than 60.
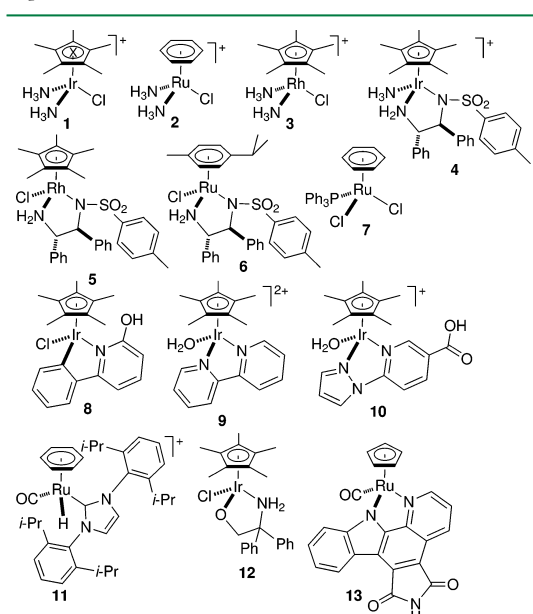
Figure 1. Representative three-legged piano stool complexes used in this study. Model compounds **1**−**3** are DFT structures, and the remaining complexes **4**−**13** are X-ray structures extracted from the CSD (**4**−**12**) or PDB (**13**).

VB[22−24] is based on valence bond theory as originally developed by Pauling.[25] It replaces the conventional harmonic bending term in a general-purpose force field with the VB-energy expression which is based on hybrid orbital strength functions as the basis for a molecular mechanics expression. These functions not only describe the energy of bond angles around the minimum but also at larger distortions where the harmonic approximation breaks down. In this fashion, VB can reproduce unusual geometries as illustrated by the distorted trigonal prism reported for $[W(CH_3)_6]$.[26] More recently, we extended the VB formalism to include the *trans*-influence which yields VB-trans (VBT).[19]

For an angle α between nonhypervalent bonds with an $sp^m d^n$ hybrid orbital, the overlap between the orbitals, Δ, is expressed as a function of the hybridizations m and n and the angle:

$$\Delta = \frac{1}{1 + m + n}\left(1 + m\cos\alpha + \frac{n}{2}(3\cos^2\alpha - 1)\right) \quad (4)$$

The overlap is used to define the strength function S:

$$S(\alpha) = S^{max}\sqrt{1 - \frac{1 - \sqrt{1 - \Delta^2}}{2}} \quad (5)$$

where S^{max} is the maximum overlap between the orbitals:

$$S^{max} = \sqrt{\frac{1}{1 + m + n}}\left(1 + \sqrt{3m} + \sqrt{5n}\right) \quad (6)$$

The energy contribution of one (bond-)orbital to the angular energy is

$$E_{angle} = k(S^{max} - S(\alpha)) \quad (7)$$

The parameter k is a scaling constant for the interactions between the atom at the center of the angle with one of its bonding partners. VB also supports hypervalent compounds using a three-center four-electron bonding model.[23,24] For transition metals, VB traditionally considers only sd hybrids. Thus complexes that count more than 12 electrons in their valence orbitals are considered hypervalent and three-center four-electron bonds are used instead of the p-orbitals.

An additional advantage of VB over conventional force fields is that it requires considerably fewer angle parameters; only one parameter is required for every pair of bonded atoms. The remaining parameters (bonds, dihedral, nonbonding, etc.) are identical to those used in the standard CHARMM force field. However, partial optimizations of these parameters are likely to be necessary for a robust force field including specific metals in their different oxidation states. In addition, assignment of a hybridization is not always trivial and needs careful exploration for more complicated cases, such as the piano stool complexes in the present study. The determination and validation of such a parametrization is the main subject of the present contribution. In our implementation of VB and VBT[19] (available from CHARMM[5] version 37 onward), the conventional CHARMM angle terms are replaced by more realistic combinations of hybrid orbitals which can be assigned by the user. In this way, VB and VBT can be combined and are compatible with the conventional CHARMM force field and simulations in a variety of ways are possible, as will be explained further below. Recently, MacKerell et al. have introduced the CHARMM general force field (CGenFF[27] which also provides a convenient fitting environment (https://www.paramchem.org) for automatically generated parameters.[28,29] We also explored the compatibility of CGenFF with VB, and the results are presented in the SI.

In the present work we apply VB for d^6-piano stool complexes which primarily involves the description of the metal interacting with the piano stool moiety. For this, a range of typical piano stool complexes is parametrized within the VB framework, and structural and dynamical properties are determined. Several parametrization strategies are pursued, compared, and validated vis-a-vis existing experimental data (structural and dynamical). However, a full parametrization of every complex, which is outside the scope of the present work, was not attempted. This applies in particular for the nonbonded

Table 1. RMSD of the crystal structure vs. the optimized structures by DFT, CHARMM and VB. RMSD5 includes only the metal, its three bonded neighbor atoms and the dummy atom. The RMSD includes all atoms except hydrogens. Individual RMSDs are given for Mulliken charges, and an overall comparison for all complexes with ChelpG and ESP charges is also reported

	metal	crystal structure	RMSD5 [Å]					RMSD [Å]				
			DFT	VB_{sp^3}	VB_{LC}	$C-VB_{LC}$	CHARMM	DFT	VB_{sp^3}	VB_{LC}	$C-VB_{LC}$	CHARMM
1	Ir		$-^a$	0.24	0.08	0.08	0.10	$-^a$	0.18	0.09	0.08	0.11
2	Ru		$-^a$	0.20	0.07	0.07	0.20	$-^a$	0.14	0.05	0.05	0.15
3	Rh		$-^a$	0.29	0.08	0.08	0.14	$-^a$	0.28	0.20	0.20	0.23
4	Ir	CSD: YOBFUR[40]	0.06	0.10	0.11	0.10	0.12	0.41	0.38	0.29	0.39	0.47
5	Rh	CSD: KIMQAY[38]	0.07	0.23	0.12	0.12	0.12	0.59	0.54	0.58	0.53	0.51
5	Rh	CSD: WOLCOP[39]	0.07	0.23	0.11	0.11	0.11	0.60	0.55	0.59	0.54	0.52
6	Ru	CSD: TAXFON[41]	0.07	0.18	0.08	0.08	0.11	0.21	0.44	0.29	0.27	0.30
7	Ru	CSD: ZESSAR[42]	0.07	0.27	0.14	0.15	0.14	0.45	0.60	0.34	0.34	0.31
8	Ir	ref 43	0.06	0.25	0.07	0.06	0.11	0.47	0.36	0.12	0.34	0.33
9	Ir	CSD: ECUBEJ[44]	0.08	0.20	0.09	0.09	0.09	0.52	0.23	0.15	0.17	0.17
10	Ir	ref 45	0.10	0.20	0.12	0.11	0.09	0.26	0.23	0.17	0.22	0.24
11	Ru	CSD: BUCLOB[46]	0.09	0.17	0.07	$-^b$	$-^b$	0.25	0.32	0.26	$-^b$	$-^b$
12	Ir	CSD: XUDMUE[47]	0.08	0.23	0.08	0.07	0.13	0.40	0.39	0.41	0.27	0.34
13	Ru	PDB: 2BZH[2]	0.07	0.12	0.09	$-^b$	$-^b$	0.11	0.29	0.16	$-^b$	$-^b$
		avg. RMSD vs X-ray[c]	0.07	0.20	0.10	0.10	0.11	0.39	0.39	0.30	0.34	0.36
		avg. RMSD vs X-ray ChelpG[c]	–	0.24	0.10	0.10	0.14	–	0.42	0.34	0.38	0.42
		avg. RMSD vs X-ray ESP[c]	–	0.26	0.11	0.11	0.17	–	0.42	0.31	0.37	0.43

^{a}The RMSD and RMSD5 was determined between the DFT and the FF structures. ^{b}The calculations were performed only using VB. ^{c}Excluding complexes **1−3**.

interactions. Rather, the work focuses on the feasibility of atomistic simulations for d[6]-piano stool complexes within a typical parametrization framework which will need to be refined for specialized applications.

COMPUTATIONAL METHODS

Molecular mechanics simulations were carried out in CHARMM[5] using the CHARMM22[30] force field with custom parameters. Further simulations were performed in CHARMM supplemented with the VB module.[19] Electronic structure calculations at the DFT[31] level were carried out with Gaussian.[32] Prototypical three-legged piano stool structures were extracted either from publications, the Cambridge Structural Database (CSD),[33] or the Protein Data Bank (PDB).[34]

Reference Compounds and Reference Calculations. With the aim of modeling important intermediates present in catalytic cycles as well as three-legged piano stool complexes acting as enzyme inhibitors, we selected $(\eta^n$-arene)Ru(II)L_3 (n = 5,6) and $(\eta^5$-Cp*)M(III)L_3, (M = Ir, Rh) complexes. A representative palette of ancillary ligands L was selected: *N*-heterocyclic carbene, phosphine, diimine, phenylpyridine, amine, aminosulfonamide, aminoalcohol, carbon monoxide, water, chloride, and hydride. All arene caps discussed in this study are planar, i.e., sp[2] hybridized. However, potential distortions from planarity in other structures can be modeled by introducing dihedral terms.

In order to derive the corresponding force field parameters, the 13 three-legged piano stool complexes were subjected to DFT calculations to compute their ground-state geometry and corresponding energy surface scans along specific internal coordinates. The studied complexes are summarized in Figure 1. Calculations were carried out using Gaussian and the hybrid B3LYP functional.[35,36] The atomic basis sets consisted of an effective core potential (LanL2DZ[37]) for the metal and the explicit 6-31G(d,p) basis for all remaining atoms. The quality of

the DFT geometries was assessed by computing the RMSD vs the experimental X-ray structures (see Table 1) where available. For complex **5**, two enantiomers were structurally determined[38,39] and thus included.

Force Fields and Parametrization. The standard procedure for deriving CHARMM-parameters uses calculations at the MP2/6-31G(d) level.[27,30] However, as calculations involving transition metals are computationally demanding, we decided to use DFT as reference calculations. One should note that irrespective of the electronic structure level chosen, parameters need to be refined by fitting to experimental data. In particular, parametrization of the nonbonded parameters may be carried out in various ways, ranging from a sequential to a fully combined fit of the nonbonded terms[48] and including experimental data of different origin.[49]

Piano Stool Moiety. In order to describe the three-legged piano stool geometry, all donor atoms of the ligands were explicitly connected to the metal using a bond. For the $(\eta^n$-arene)M interaction (n = 5,6) a dummy atom X at the barycenter of the ring was introduced.[50] It was incorporated in CHARMM and VB using the lonepair module and thus kept at the center of the ring throughout the simulation. In order to maintain an sp[2] hybridization of the arene carbons, no explicit bond between the dummy atom and the arene carbons was introduced. To account for the tilting of the ring, M−X−C_{arene} angles were included. The M−C_{arene} van der Waals (vdW) and electrostatic interactions were excluded to avoid redundant counting of interactions already present in the M−X bond. Dihedral terms in metal complexes are generally neglected[18] which was also done here. The following force field combinations were considered in more detail.

Pure VB:VB_{sp^3} Model. According to VB, all piano stool complexes would be hypervalent. However, by introducing a dummy atom as a point of reference for the arene, the hypervalency concept no longer applies, as the bond between the dummy atom and the metal accounts for more than two

107

electrons. Thus we decided to work within nonhypervalent VB by using $sp^m d^n$ hybridizations. In light of the $(\eta^n\text{-arene})M(L)_3$ structure, reminiscent of a tetrahedron, the tetrahedral sp^3 hybridization was shown to be a good starting point to optimize the geometry of the three-legged piano stool.

Pure VB:VB$_{LC}$ Model. As angles observed in X-ray structures tend to cluster around 90° (between the piano stool legs) or 130° (between the ring-center and the piano stool legs), the tetrahedral approximation with its canonical 109.5° angles is not optimal. Instead, a linear combination of other hybridizations can be envisaged. As shown in Figure 2, the angle-
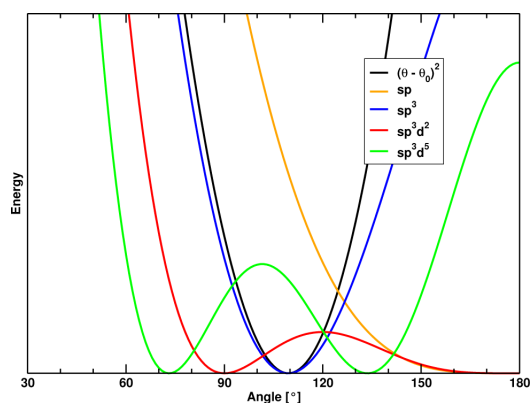
Figure 2. VB energy (no scale given) as a function of the angle for several hybridizations. For comparison, the energy function of a harmonic angle with $\theta_0 = 109.5°$ is shown as well.

dependent VB-energy for $sp^3 d^2$ has a minimum at 90°, whereas for $sp^3 d^5$ its minimum is at 130°. Therefore $sp^3 d^2$ was used as hybridization of the piano stool leg orbitals, whereas the hybridization for the M–X orbital was $sp^3 d^5$. Such orbitals have been used to describe OC–M–CO angles previously,[51] but their use here is empirical, as a combination $sp^3 d^2$ and $sp^3 d^5$ hybridizations is used on the metal.

Standard CHARMM. All angles are treated with standard CHARMM, no VB is in use.

Mixed CHARMM/VB:C-VB$_{LC}$ Model. A mixed VB and CHARMM potential energy function whereby a VB-energy term is combined with conventional CHARMM angle terms. Only the X–M–L and L–M–L angles surrounding the metal are treated by VB, whereas all other angles are calculated with CHARMM.

Parametrization. Force field parametrization was carried out by fitting the parameters to reproduce the energy profiles of energy surface scans obtained from DFT calculations. Parameters not available in the standard CHARMM22 force field were those either derived previously[52] or taken from parameters for related chemistries through analogies. The fitting was performed using the full force field as opposed fitting the respective parameter without other terms involved. To assist parametrization, perl scripts from PerlMol (www.perlmol. org) were used to generate the topology and necessary input files for the CHARMM calculations directly from the Gaussian output files. The fitting itself was performed using Chnolls,[49] the CHARMM interface for Inolls.[53] For the charges of the ligand atoms, Mulliken charges from the DFT optimizations were used and not specifically optimized. To assess the

influence of the charge scheme used for these complexes, we also tested ChelpG[54] and ESP[55,56] charges, and the results are reported in Tables 1 and S3. It is worthwhile to mention that full compatibility with CHARMM would require fitting nonbonded parameters to interaction energies with a water molecule. No such full parametrization was attempted in the present case which, however, would be certainly possible and necessary for specific applications. vdW parameters for all atoms except for the metal were those from the CHARMM22 force field, whereas for the metal atoms those from UFF[57] were used. As CHARMM and UFF use different mixing rules[58] this could certainly be improved. However, for all complexes studied here, the metal is buried by the ligands and vdW contact with, e.g., the surrounding solvent is minor (see also Figures 8 and 9). A flowchart summarizing the parametrization is presented in Figure 3. To validate the procedure, the
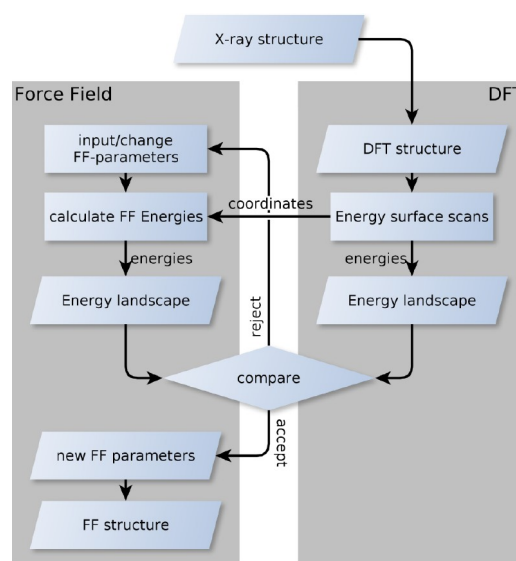
Figure 3. Summary of the force field parametrization procedure.

bidentate ligands present in complexes **6** and **9** were also parametrized by using CGenFF[27] for the ligands within models VB$_{LC}$ and VB$_{sp^3}$ models. The corresponding results are reported in Table S3.

All VB parameters were set to the default[22−24] ones except for the scaling parameter k_{Ru-N} which is 100 kcal/mol for VB$_{sp^3}$. For the more comprehensive VB$_{LC}$ model, the scaling parameters $k_{metal-ligand}$ were refitted where necessary. All bond and angle parameters were fitted or assigned, and only a small number of dihedral terms was required to obtain satisfactory results. The recommended CHARMM22 cutoffs and settings[30] were used in the parametrizations and simulations: a switching (vdW terms) respectively shifting (electrostatics) function between 10 and 12 Å, cutoffs at 12 Å, and atom-based electrostatics.

Force Field Calculations. *Structure Optimization.* Of the 13 complexes that were considered, 10 have been structurally characterized by experiment. First, the structures of all complexes were optimized with DFT and with CHARMM and VB using the conjugate gradient method until convergence
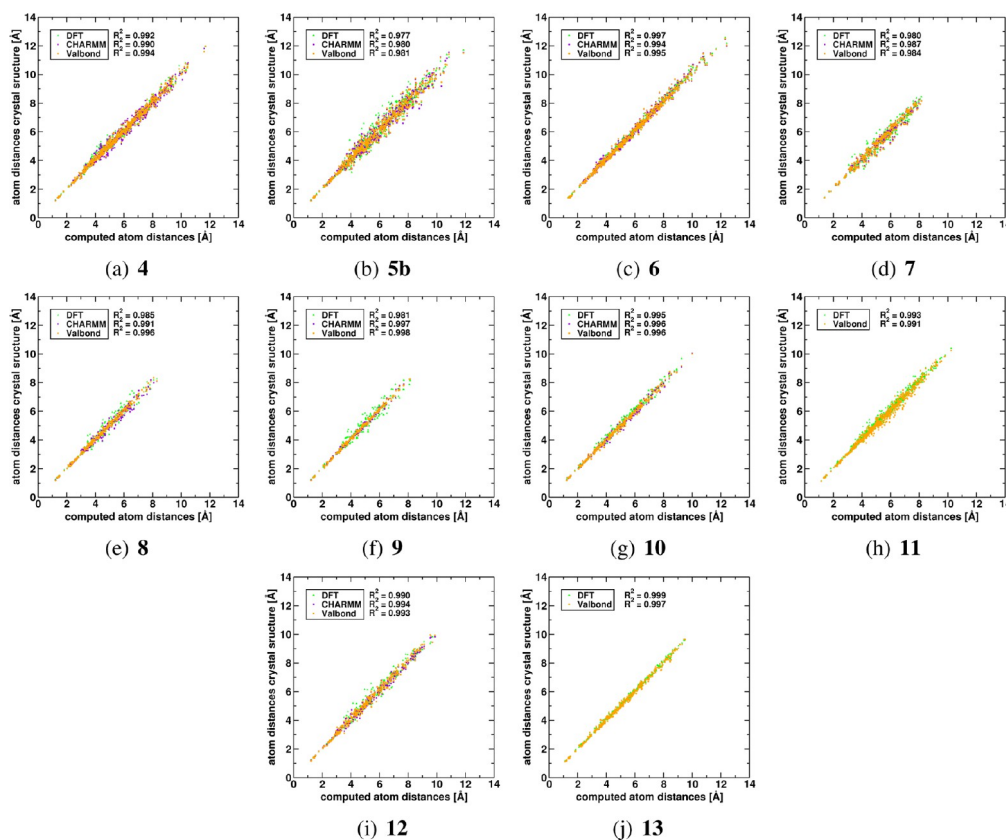
Figure 4. Comparison of the atom distances in the crystal (*y*-axis) and the calculated (*x*-axis) structures. Values for the distances in the DFT-optimized structures are displayed in green, for the CHARMM optimized structures in purple and in the VB optimized structures (VB$_{LC}$ model) in orange.

($\Delta E < 10^{-5}$ kcal/mol) was achieved. The structures were then compared by calculating the RMSD between X-ray and the energy minimized structures for all atoms except hydrogens (RMSD) and for the five "core atoms" (i.e., the three donor atoms, the dummy atom, and the metal, RMSD5), respectively. In addition, the complete distance matrix between pairs of heavy atoms was determined from the reference and computationally optimized structures.

MD Simulation in the Gas Phase. For selected complexes, molecular dynamics (MD) simulations with each of the computational models (VB$_{sp^3}$, VB$_{LC}$, CHARMM, and C-VB$_{LC}$) were performed. An initial 10 ps (with a stepsize of $\Delta t = 0.2$ fs) heating period from 100 to 300 K was followed by 20 ps of equilibration at 300 K. No SHAKE[59] was used. Following this, 2 ns trajectories in the NVE (microcanonical) ensemble were run to assess the stability of the simulations. All simulation conditions were identical to those used for the minimizations. As demonstrated previously,[60] VB simulations can be carried out on the 100 ns time scale. However, for the purpose of the present work (validation of energy conservation), this was not deemed necessary.

Explicit Solvent Simulations. For molecule **1**, MD simulations with CHARMM in explicit solvent were performed in two different solvents. The VB$_{LC}$ approach in combination

with standard CHARMM was chosen to perform these simulations to illustrate the interplay of the VB force field with conventional CHARMM. VB was only used for the metal complex, and the solvents were treated with conventional CHARMM. The first simulation was carried out in explicit water (TIP3P model).[61] The second simulation was performed in a methanol solvent box using parameters from the CHARMM22 force field. The simulations were done with periodic boundary conditions in a cubic solvent box of 32 Å (methanol) or 35 Å (water) edge length. A time step of 1 fs was used. SHAKE[59] was employed on the bonds involving hydrogen atoms. After heating and equilibration, simulations were done in the NPT ensemble, keeping pressure and temperature constant with the extended system thermostat.[62−64] A total of 5 ns were simulated for the complex with each of the solvents. Cutoffs and remaining settings were identical as above.

NMR Calculations. To further validate the parametrizations against solution-phase data, chemical shifts for complex **9** were determined. Two simulations in explicit water for complex **9** were carried out: one using the same conditions as above and a second simulation without employing SHAKE for the hydrogens of the metal complex. The ^{1}H and ^{13}C chemical shifts were calculated for both the DFT-optimized structures and the

109

Article

500 snapshots from the MD simulations using the gauge-independent atomic orbital (GIAO) method.[65] From the 500 snapshots, a "spectrum" was generated by calculating the distribution of the chemical shifts in the snapshots. Two functionals were used, namely the standard B3LYP and the dispersion corrected M06-2X functionals.[66]

■ RESULTS AND DISCUSSION

Reference Compounds, Reference Calculations, and Parametrization. The largest RMSD between the X-ray and the DFT-optimized structures for the five "core atoms" (RMSD5) is 0.10 Å (see Table 1), demonstrating that DFT calculations are well suited to reproduce ground-state piano stool geometries. The influence of the charge models used in the parametrization (Mulliken, ChelpG, and ESP) on describing the structural data was found to be rather small. The RMSD5 is virtually the same for all charge models, the RMSD is slightly smaller with the Mulliken charges. Typically, the charges differed by less than 0.1 e. But for the metal charges, deviations of up to 0.4 e were observed. The distance matrices (Figure 4) correlate well with X-ray geometries ($R^2 = 0.98$–0.99).

In a next step, the bonded parameters (bond, angles) of the force field were parametrized. For this purpose, energy surface scans were carried out for complexes **1**–**3**, **7**, and **10**–**12** along different relevant internal coordinates. The bonds scanned are summarized in Table 2, and the list of scanned angles is

Table 2. M−L Bond Length Parameters and VB Orbital Scaling Parameters[a]

atoms		bond parameters		VB parameter	fitting on structure
		force constant kcal/mol/Å²	equilibrium Å	scaling factor[b] kcal/mol	
Ru	X	210	1.72	140	**2**
Ru	N	113	2.28	500	**2**
Ru	Cl	105	2.50	350	**2**
Ru	$C_{carbene}$	106	2.05	100[c]	**11**
Ru	C_{CO}	206	1.80	100[c]	**11**
Ru	H	138	1.54	100[c]	**11**
Ru	P	62	2.08	100[c]	**7**
Rh	X	210	1.81	100[d]	**3**
Rh	N	70	2.38	285[d]	**3**
Rh	Cl	77	2.54	260[d]	**3**
Ir	X	210	1.81	100	**1**
Ir	N	100	2.30	285	**1**
Ir	Cl	113	2.50	260	**1**
Ir	O	145	2.11	250	**12**
Ir	C	130	2.00	100[c]	**10**

[a]Equilibrium parameters are not necessarily identical to the bond length found in X-ray structures, as the parameters are fitted to reproduce energy surface scans and not minimum-energy structures. [b]Only for the VB_{LC} model, default[22−24] values for VB_{sp^3} model. [c]Default value, not fitted. [d]Fitted for iridium.

collected in the SI. The force field parameters were optimized by using the Chnolls platform as described above to minimize the difference in energy between the force field and DFT energy landscapes. Relevant optimized bond parameters and refitted VB scaling factors are collected in Table 2. A complete list of all parameters in CHARMM format is collected in the SI as are the topology files for all complexes.

For the VB_{sp^3} model, only one VB parameter (k_{Ru-N}) needed to be adapted for accurately reproducing the minimum-energy structures for the 13 molecules. All other VB parameters could be kept at their default values, and no further adjustments were necessary, thus considerably reducing the amount of reparametrization required. For the more elaborate model with the linear combination of hybridizations (VB_{LC}), seven parameters were refitted. In comparison, more than 200 nonstandard CHARMM angle parameters were parametrized or assigned by analogy, even though we reduced the test set for CHARMM to only 11 from the original 13 molecules. Including complexes **11** and **13** would require an additional 40 terms to be determined.

Force Field Calculations. *Structure Optimization.* In order to estimate the quality of the force field-calculated structures for complexes **4**−**13**, RMSD and RMSD5 relative to the X-ray structures were calculated. For model complexes where no X-ray data are available (complexes **1**−**3**), the RMSD were computed relative to the DFT-optimized structures. RMSD5 for the VB_{sp^3} model is slightly higher compared to the DFT optimized structures (Table 1). In contrast, the RMSD for the entire structures (excluding hydrogens) is comparable or even smaller than that for the DFT structures. The average RMSD vs X-ray is the same for DFT and the VB_{sp^3} model (0.39 Å), whereas the average RMSD5 for the VB_{LC} model, the CHARMM-only parametrization and the C-VB_{LC} model is virtually the same. The average global RMSD for these three cases is slightly lower than for DFT, with VB_{LC} performing best. As expected, the results for the two enantiomers of complex **5** are very similar. Distance matrices show a similar trend (Figure 4) with deviations for the force fields slightly smaller than for the DFT calculations. For comparison, all superimposed structures are presented in Figure 5. As can be appreciated, the force field and DFT-calculated structures are strikingly similar to the X-ray structures. The largest deviations arise from rotations of aromatic residues in the second coordination sphere of the metal. Crystal contacts in the X-ray structures might be responsible for this, and improvements may be possible by introducing additional dihedral angle parameters. The VB optimized structures for complexes **11** and **13** yield a nonlinear M−C−O arrangement ($\alpha = 160°$) with tilting of the C−O toward the η^n arene moiety. This may be a limitation in the VB approach as the energy surface for sp hybridizations is very flat (see Figure 2), and thus the distortion of the angle from 180° is not hindered enough by the force field. This issue was resolved by adding a small (10 kcal/mol) harmonic constraint around 180°.

A 72° rotation of the Cp* ring results in a structure which is energetically and structurally indistinguishable (Figure 6). To evaluate the torsional barrier of the Cp* rotation, a potential surface scan for the rotation along the Cp*−Ir bond was performed on complex **1** at the DFT level. The maximal energy variation was found to be <0.2 kcal/mol with a quasiperiodicity of 72° (Figure 6).

The four models differ in number and types of parameters needed. VB_{sp^3} is straightforward and required least parametrization. The results are robust and valid for exploratory simulations, although with the present parametrizations it shows the largest deviations. The VB_{LC} model required reparametrization of several VB parameters, and the results are meaningful. The CHARMM approach involves most parametrization, and the results are slightly worse for the angles involving metals. Using the C-VB_{LC} model is a compromise between the CHARMM and the VB force field,

F

dx.doi.org/10.1021/ct301080d | *J. Chem. Theory Comput.* XXXX, XXX, XXX−XXX

110
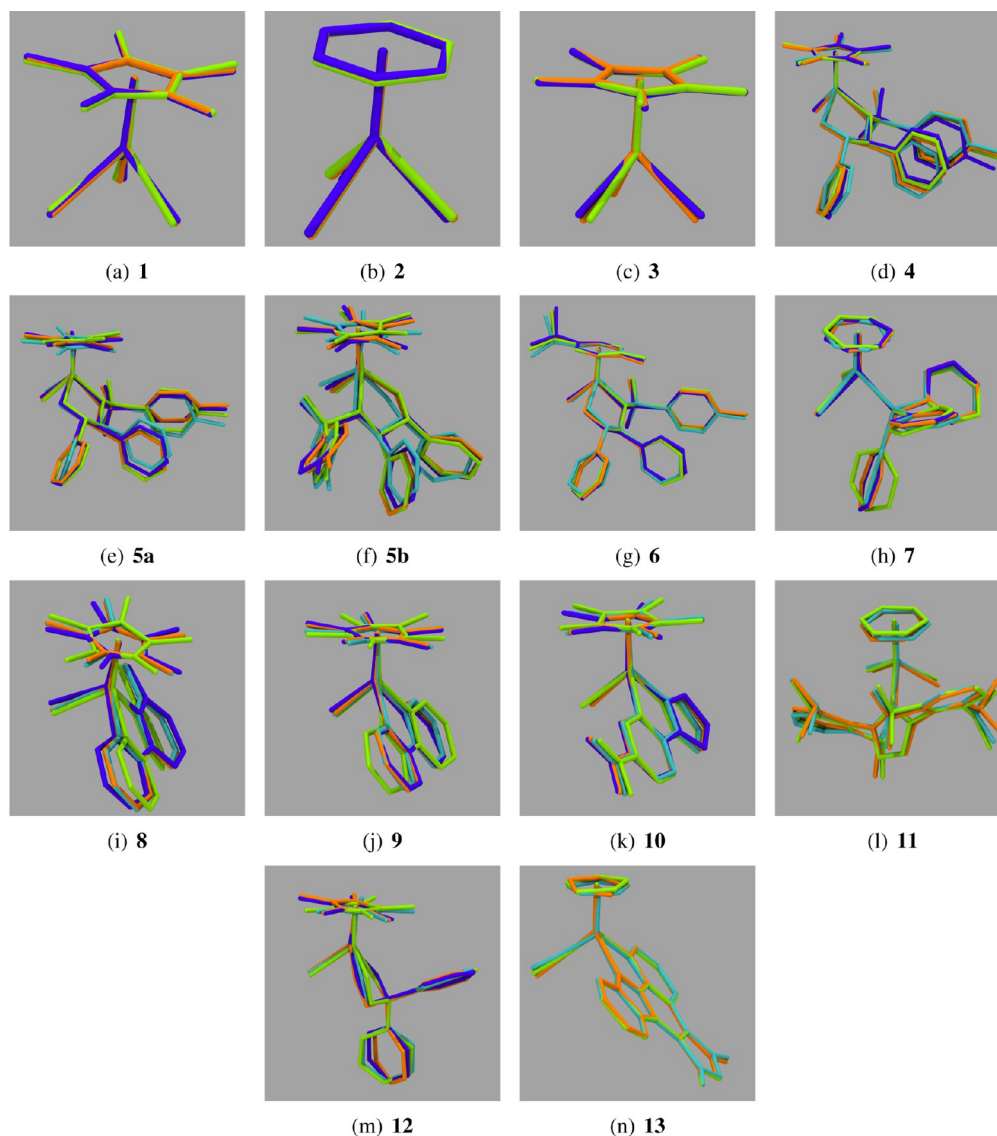
Figure 5. Superposition of the crystal structure (cyan) with the DFT-optimized (green), the CHARMM optimized (purple) and the VB optimized (VB$_{LC}$ model, orange) structures. The superposition minimizes the RMSD.

and VB is only used for the L−M−L angles. As the L−M−L angles are best described by VB and given that reliable CHARMM parameters for the ligands are available, e.g., through parametrization with CGenFF, we recommend the C-VB$_{LC}$ as the default approach for production simulations. If no parameters are readily available, we recommend the VB$_{LC}$ model. For all these models it should be emphasized that depending on the purpose, further reparametrization in particular of the nonbonded parameters may be required. Also, the merits and disadvantages of the four models explored here may be more or less accentuated if such a more complete parametrization is carried out.

MD in Gas Phase. To verify that the forces were correctly implemented and to demonstrate the capability of following explicit dynamics of a system in a representative environment, MD simulations were performed. Because the resulting trajectories all display similar behavior of the complexes, we focused on simulations of molecules **1** and **4** using the VB$_{LC}$ approach for a more detailed analysis.

The total energy is conserved during the 2 ns NVE VB MD simulation and the temperature as well as the kinetic and potential energies are Gaussian distributed as expected, see Figure S4. For the RMSD of all atoms except hydrogens (RMSD), several stable states can be identified (see Figure 7). Each of these corresponds to a 72° rotation of the *Cp** moiety
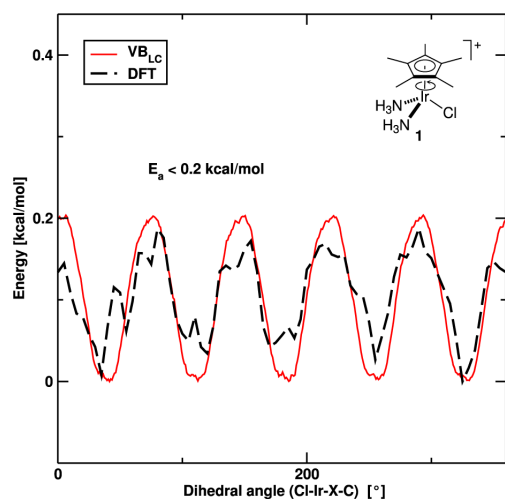
Figure 6. Computed energy profile for the rotation of $Cp*$ around the X−Ir axis in complex **1**. (Black dashed line is the DFT energy. Red line is the energy according to probability distribution of the dihedral angle for the MD simulation with the VB_{LC} model.).

along the X−Ir axis and illustrates the low activation energy required for an M−$Cp*$ rotation. The probability distribution of the dihedral angle Cl−Ir−X−C results in an energy barrier of 1.5 kcal/mol for complex **4**. For complex **1** it yields 0.2 kcal/mol, reproducing ΔE calculated using DFT with remarkable accuracy (see Figure 6). According to transition-state theory, $1/\tau_{A \to B} = (k_b T/h) \exp(-\Delta G_{A \to B}/k_b T)$, the states have lifetimes on the ps time scale which can be confirmed by visual analysis of the rotation where the lifetime of the states is ≈ 10 ps for complex **4** and ≈ 1 ps for complex **1**, respectively.

At $t = 1.6$ ns one of the phenyl rings of the sulfonamide ligands flips, resulting in the observed increase of the RMSD. The RMSD5 remains <0.3 Å, and no drift can be observed. For several hundred snapshots energy calculations using DFT and the C-VB_{LC} force field have been carried out. The correlation between force field and DFT calculations is $R = 0.8$. This is an acceptable correlation but could be further improved by refitting the nonbonded parameters which, however, is outside the scope of the present work. The results demonstrate that robust NVE simulations are possible which opens the way for more elaborate investigations. As reported in Table S3, the results from calculations based on the CGenFF parametrizations were similar to simulations using our parameters (slightly worse results for complex **9** and slightly better results for complex **6**), which illustrates the robustness of the present approach.

Explicit Solvent Simulations. For organometallic complexes, the dynamics in a realistic solvent environment is of particular interest because additional factors influencing their activity and behavior arise. They include direct solvent coordination, electrostatic shielding, and entropic effects and can affect conformations and reactivities of the chemical species involved. Often, such effects are excluded or only taken into account in a highly averaged fashion in computational investigations of transition-metal compounds because a realistic solvent environment increases the number of atoms by at least an order of magnitude and renders such studies impractical. In the following, we consider the organization of solvent molecules around complex **1** in two different solvents as an example.

The RMSD over the duration of the simulations in explicit solvent behave similarly to simulations in the gas phase. Simulations in water or methanol do not significantly differ. To investigate the effects of the solvent on the simulation, i.e., the solvation shell, the solvent distribution around the metal complex during the simulation was characterized. The solvent is represented as a density map for each element type, see Figures
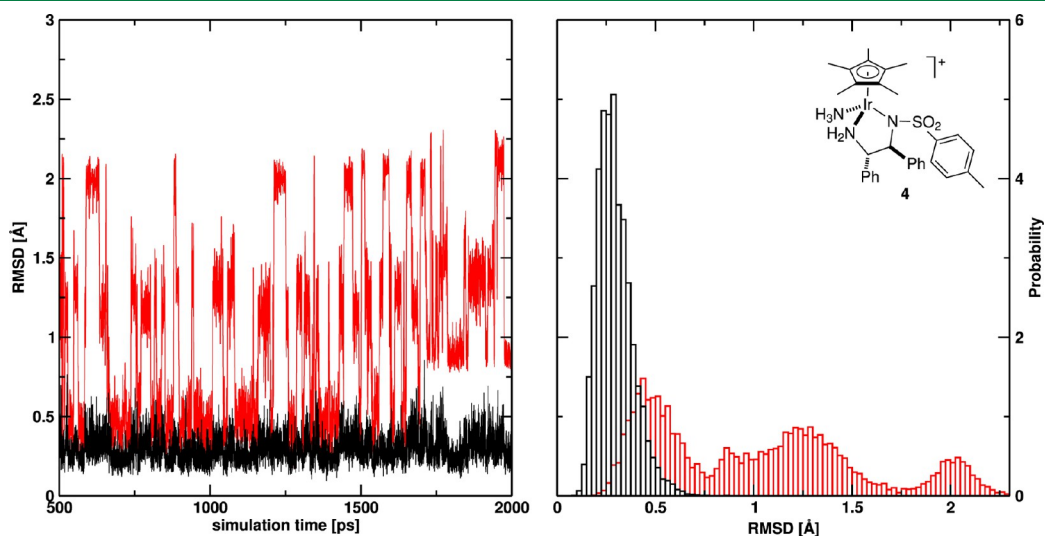
Figure 7. RMSD during the MD simulation using VB_{LC} for complex **4**. For the 5 core atoms (black traces), the RMSD remains <0.5 Å during most of the simulation. Inclusion of all heavy atoms (red traces) reveals higher deviations diagnostic of the rotation of the $Cp*$ moiety along the M−$Cp*$ axis.

(a)   (b) **Oxygen**   (c) **Hydrogen**
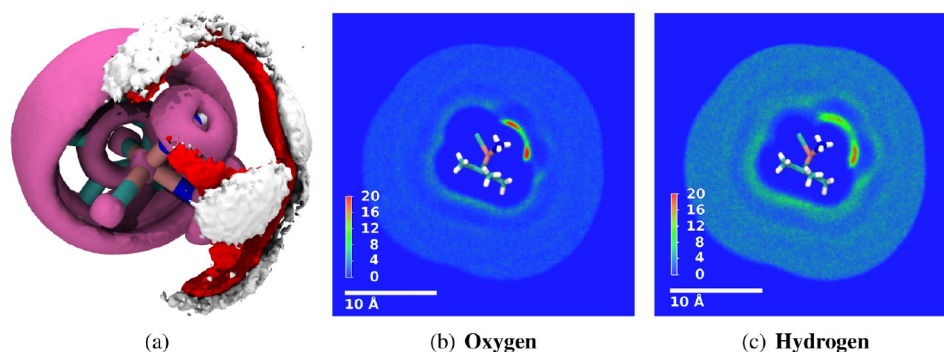
Figure 8. Solvent distribution for an MD simulation of complex **1** in explicit water. (a) 3D representation of the atom distribution around the metal complex during the simulation. The probability density for the piano stool atoms is colored in magenta, whereas for the solvent atoms it is depicted as separate densities for each element (oxygens, red; hydrogens, white). The isosurfaces are drawn at 5%, 30%, and 40% of the maximum density for the metal complex, the water–oxygens and the water–hydrogens, respectively. (b) and (c) Slice through the probability distribution for oxygens (b) and hydrogens respectively (c) for the duration of the simulation with a cutoff of 10 Å around the complex. The slice was cut through the plane formed by the metal, the dummy atom, and the chloride. The scale in (b) and (c) reports total counts in voxels of size $0.2 \times 0.2 \times 0.2$ Å^3.

8a and 9a. Each snapshot is oriented such as to minimize RMSD5 and thus remove the translational and rotational degrees of freedom.

The resulting density for the metal-complex atoms shows the rotation of the Cp* ring for both solvents. It also reveals only very little movement of the metal-connecting atoms. Slices through these density maps are displayed in Figures 8 and 9 for water and methanol, respectively. The density was cut through the mirror plane formed by the M–X–Cl plane, slicing the complex into two mirror images. The oxygen-atom density in both cases (water and methanol) is highest next to the ammonia hydrogens, a result of the polar interactions. A first and a second solvation shell with a region of lower probability in between can clearly be identified. The carbon density is highest on top of the Cp* ring highlighting the hydrophobicity of the Cp* moiety. The high densities close to the ammonia groups result from orientation of the methanol molecules to present the oxygen toward the hydrogens. The hydrogen atom densities concur with the observations made for the carbon and oxygen atoms. In regions where the heavy-atom densities are highest, the solvent molecules are more ordered, and thus the hydrogen densities are more localized as well. Overall, ordering is dominated by hydrophilic rather than hydrophobic interactions as the respective solvent densities are more influenced by the higher charged nitrogen than the less charged carbon atoms of the ligand.

The configurations sampled can also be used to further assess the force field parametrization. We stress again, that no attempt was made to further refine the nonbonded parameters to optimally reproduce reference electronic structure calculations, as such quantitative studies are not the focus of the present work. This can in principle done using available tools.[49] Rather, an ensemble of 1000 complex:solvent structures including the complex with **1** and **3** nearby water molecules were extracted from the simulations and their stabilization energies were computed from VB$_{LC}$ and DFT (B3LYP and M06-2X functional with basis sets as described above) electronic structure calculations. A graphical comparison of the two methods is provided in Figure S5, and the correlation coefficients are $R = 0.62$ and 0.66, respectively.

NMR Calculations. In order to assess the suitability of such simulations in solution, NMR chemical shifts from explicitly solvated simulations in H$_2$O were determined for complex **9**. Figure S6 compares the experimentally determined shifts[44] with those computed for the DFT-optimized structures and the conformationally averaged shifts. The experimental values were determined in D$_2$O at 298 K. For the ^{1}H chemical shifts, surprisingly, the DFT B3LYP calculation performs better than the dispersion-corrected M06-2X functional. MD simulations lead to significant broadening of the peaks due to conformational sampling. This is particularly evident for simulations without SHAKE as the hydrogen bonds are allowed to fluctuate. However, the maxima of the distributions match the experimentally determined chemical shifts quite well. A recent parametrization of sulfonyl containing compounds[67] also used NMR data but in a somewhat more qualitative fashion to validate the parametrizations. In line with the present work, the distributions of the geometrical coordinates used to make contact with the experimentally observed chemical shifts are broad. The present comparison suggests that the force field parameters are a useful starting point but also that there is room for improvement should the particular application require that.

■ CONCLUSIONS

Parameters, suitable for atomistic simulations of d^6 three-legged piano stool complexes with the VB and the CHARMM force field, have been developed and tested for minimum-energy structures and explicit simulations. Specifically, (η^n-arene)ML$_3$ ($n = 5,6$) complexes with ruthenium, rhodium and iridium metals and with N-heterocyclic carbene, phosphine, diimine, phenylpyridine, amine, aminosulfonamide, aminoalcohol, carbon monoxide, water, chloride, and hydride as ligands were parametrized. With these parametrizations the complexes can be treated either using the traditional CHARMM or the VB extended CHARMM force fields. From a parametrization perspective, the VB force field is advantageous, as it requires fewer parameters. Specifically, only bonded parameters need to be parametrized, the default VB angle parameters[22−24] with Mulliken charges and vdW parameters from the UFF force field yield meaningful results thus considerably reducing the

Article
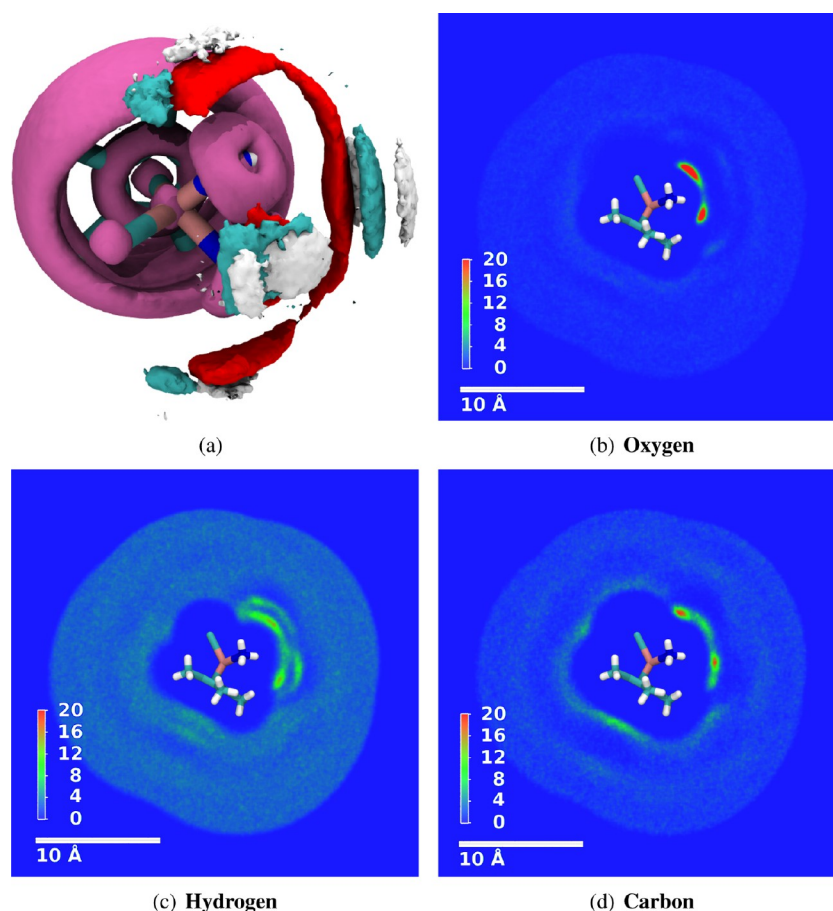
Figure 9. Solvent distribution for an MD simulation of complex **1** including explicit methanol. (a) 3D representation of the atom distribution around the metal complex during the simulation. The probability density for the piano stool atoms is colored in magenta, whereas for the solvent atoms it is depicted as separate densities for each element (oxygens, red; hydrogens, white; and carbons, cyan). The isosurfaces are drawn at 5%, 30%, and 40% of the maximum density for the metal complex, the water−oxygens and the water−hydrogens, respectively. (b−d) Slice through the probability distribution for oxygens (b), hydrogens (c), and carbons (d), respectively, for the duration of the simulation with a cutoff of 10 Å around the complex. The slice was cut through the plane formed by the metal, the dummy atom, and the chloride. The scale in (b−d) reports total counts in voxels of size $0.2 \times 0.2 \times 0.2$ Å^3.

parametrization effort. As VB is compatible with the traditional CHARMM angular parameters, mixed VB/CHARMM calculations can be performed to combine the strengths of both approaches: the established quality of CHARMM for simulations of large biomolecules and the versatility of VB for unusual geometries of metal complexes.

The present work establishes that geometrical structures can be reliably reproduced compared to X-ray and DFT structures, respectively. The RMSDs between X-ray and structures for the four models are comparable to RMSDs obtained between X-ray and DFT structures. Extensive MD simulations are possible which allows, in principle, to investigate the solvent structure around the organometallic complex and capture effects, such as entropy changes upon binding. However, for specific catalysts and depending on the concrete application, further refinement, e.g., by fitting to vibrational or thermodynamic data, of the force field will be necessary. The techniques presented here will be useful starting points for investigating catalysts in dynamically more demanding environments, including their roles in artificial metalloenzymes and as enzyme inhibitors.

■ **ASSOCIATED CONTENT**

Ⓢ **Supporting Information**

Force field parameters, topology files, an example input file, additional tables (list of scanned L−M−L angles, metal distances and comparison to CGenFF) and figures (describing the fitted PES, metal distances, the correlation between DFT and FF energies for complexes with water, a comparison of calculated and experimental chemical shifts, and energy distribution during the simulation). This material is available free of charge via the Internet at http://pubs.acs.org.

■ **AUTHOR INFORMATION**

Corresponding Author
*E-mail: thomas.ward@unibas.ch; m.meuwly@unibas.ch

114

Notes

The authors declare no competing financial interest.

REFERENCES

(1) Murata, K.; Ikariya, T.; Noyori, R. *J. Org. Chem.* **1999**, *64*, 2186–2187.
(2) Debreczeni, J. E.; Bullock, A. N.; Atilla, G. E.; Williams, D. S.; Bregman, H.; Knapp, S.; Meggers, E. *Angew. Chem., Int. Ed.* **2006**, *45*, 1580–1585.
(3) Zimbron, J. M.; Sardo, A.; Heinisch, T.; Wohlschlager, T.; Gradinaru, J.; Massa, C.; Schirmer, T.; Creus, M.; Ward, T. R. *Chem.—Eur. J.* **2010**, *16*, 12883–12889.
(4) Monnard, F. W.; Heinisch, T.; Nogueira, E. S.; Schirmer, T.; Ward, T. R. *Chem. Commun.* **2011**, *47*, 8238–8240.
(5) Brooks, B. R.; Bruccoleri, R. E.; Olafson, D. J.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187–217.
(6) Case, D. A. et al. *Amber 12*; University of California, San Francisco: San Francisco, CA, 2012.
(7) Schuler, L. D.; Daura, X.; van Gunsteren, W. F. *J. Comput. Chem.* **2001**, *22*, 1205–1218.
(8) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J. Comput. Chem.* **1996**, *118*, 11225–11236.
(9) Bosnich, B. *Chem. Soc. Rev.* **1994**, *23*, 387–395.
(10) Polzer, T.; Kiefer, W. *J. Organomet. Chem.* **1996**, *508*, 153–157.
(11) de Hatten, X.; Cournia, X.; Huc, I.; Smith, J.; Metzler-Nolte, N. *Chem.—Eur. J.* **2007**, *13*, 8139–8152.
(12) Bernhardt, P. V.; Comba, P. *Inorg. Chem.* **1992**, *31*, 2638–2644.
(13) Norrby, P.-O.; Brandt, P. *Coord. Chem. Rev.* **2001**, *212*, 79–109.
(14) Piquemal, J.-P.; Williams-Hubbard, B.; Fey, N.; Deeth, R. J.; Gresh, N.; Giessner-Prettre, C. *J. Comput. Chem.* **2003**, *24*, 1963–1670.
(15) Comba, P.; Hambley, T. W.; Martin, B. *Molecular Modeling of Inorganic Compounds*; Wiley-VCH: Weinheim, Germany, 2009; p 344.
(16) Zimmer, M. *Coord. Chem. Rev.* **2009**, *253*, 817–826.
(17) Deeth, R. J.; Anastasi, A.; Diedrich, C.; Randell, K. *Coord. Chem. Rev.* **2009**, *253*, 795–816.
(18) Hu, L.; Ryde, U. *J. Chem. Theory Comput.* **2011**, *7*, 2452–2463.
(19) Tubert-Brohman, I.; Schmid, M.; Meuwly, M. *J. Chem. Theory Comput.* **2009**, *5*, 530–539.
(20) Boeyens, J. C.; Comba, P. *Coord. Chem. Rev.* **2001**, *212*, 3–10.
(21) Comba, P.; Zimmer, M. *J. Chem. Educ.* **1996**, *73*, 108–110.
(22) Root, D. M.; Landis, C. R.; Cleveland, T. *J. Am. Chem. Soc.* **1993**, *115*, 4201–4209.
(23) Cleveland, T.; Landis, C. R. *J. Am. Chem. Soc.* **1996**, *118*, 6020–6030.
(24) Landis, C. R.; Cleveland, T.; Firman, T. K. *J. Am. Chem. Soc.* **1998**, *120*, 2641–2649.
(25) Pauling, L. *J. Am. Chem. Soc.* **1931**, *53*, 1367–1400.
(26) Landis, C. R.; Cleveland, T.; Firman, T. K. *J. Am. Chem. Soc.* **1995**, *117*, 1859–1860.
(27) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; Mackerell, A. D. *J. Comput. Chem.* **2010**, *31*, 671–690.
(28) Vanommeslaeghe, K.; MacKerell, a. D. *J. Chem. Inf. Model.* **2012**, *52*, 3144–3154.
(29) Vanommeslaeghe, K.; Raman, E. P.; MacKerell, a. D. *J. Chem. Inf. Model* **2012**, *52*, 3155–3168.
(30) MacKerell, A. D.; et al. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.
(31) Hohenberg, P.; Kohn, W. *Phys. Rev.* **1964**, *136*, B864–B871.
(32) Frisch, M. J. et al. *Gaussian 03*, revision B.01; Gaussian, Inc.: Wallingford, CT, 2006.
(33) Allen, F. H. *Acta Crystallogr., Sect. B: Struct. Sci* **2002**, *58*, 380–388.
(34) Bernstein, F. C.; Koetzle, T. F.; Williams, G. J.; Meyer, E. F.; Brice, M. D.; Rodgers, J. R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. *Arch. Biochem. Biophys.* **1978**, *185*, 584–591.
(35) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648–5652.
(36) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. *J. Phys. Chem.* **1994**, *98*, 11623–11627.
(37) Hay, P. J.; Wadt, W. R. *J. Chem. Phys.* **1985**, *82*, 270–284.
(38) Mashima, K.; Abe, T.; Tani, K. *Chem. Lett.* **1998**, 1199–1200.
(39) Mao, J.; Baker, D. C. *Org. Lett.* **1999**, *1*, 841–843.
(40) Heiden, Z. M.; Gorecki, B. J.; Rauchfuss, T. B. *Organometallics* **2008**, *27*, 1542–1549.
(41) Uematsu, N.; Fujii, A.; Hashiguchi, S.; Ikariya, T.; Noyori, R. *J. Am. Chem. Soc.* **1996**, *118*, 4916–4917.
(42) Elsegood, M. R.; Tocher, D. A. *Polyhedron* **1995**, *14*, 3147–3156.
(43) Fujita, K.-i.; Yoshida, T.; Imori, Y.; Yamaguchi, R. *Org. Lett.* **2011**, *13*, 2278–2281.
(44) Ogo, S.; Makihara, N.; Kaneko, Y.; Watanabe, Y. *Organometallics* **2001**, *20*, 4903–4910.
(45) Maenaka, Y.; Suenobu, T.; Fukuzumi, S. *J. Am. Chem. Soc.* **2012**, *134*, 367–374.
(46) Saker, O.; Mahon, M. F.; Warren, J. E.; Whittlesey, M. K. *Organometallics* **2009**, *28*, 1976–1979.
(47) Suzuki, T.; Morita, K.; Tsuchida, M.; Hiroi, K. *Org. Lett.* **2002**, *4*, 2361–2363.
(48) Kramer, C.; Gedeck, P.; Meuwly, M. *J. Chem. Theory Comput.* **2013**.
(49) Devereux, M.; Meuwly, M. *J. Chem. Inf. Model.* **2010**, *50*, 349–357.
(50) Doman, T. N.; Landis, C. R.; Bosnich, B. *J. Am. Chem. Soc.* **1992**, *114*, 7264–7272.
(51) Pauling, L. *Proc. Natl. Acad. Sci. U.S.A.* **1978**, *75*, 569–572.
(52) Schmid, M.; Nogueira, E. S.; Monnard, F. W.; Ward, T. R.; Meuwly, M. *Chem. Sci.* **2012**, *3*, 690.
(53) Law, M.; Hutson, J. *Comput. Phys. Commun.* **1997**, *102*, 252–268.
(54) Breneman, C. M.; Wiberg, K. B. *J. Comput. Chem.* **1990**, *11*, 361–373.
(55) Singh, U. C.; Kollman, P. A. *J. Comput. Chem.* **1984**, *5*, 129–145.
(56) Besler, B. H.; Merz, K. M.; Kollman, P. A. *J. Comput. Chem.* **1990**, *11*, 431–439.
(57) Rappe, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard, W. A.; Skiff, W. M. *J. Am. Chem. Soc.* **1992**, *114*, 10024–10035.
(58) Halgren, T. a. *J. Am. Chem. Soc.* **1992**, *114*, 7827–7843.
(59) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327–341.
(60) Huang, J.; Häussinger, D.; Gellrich, U.; Seiche, W.; Breit, B.; Meuwly, M. *J. Phys. Chem. B* **2012**, *116*, 14406–14415.
(61) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–936.
(62) Andersen, H. C. *J. Chem. Phys.* **1980**, *72*, 2384–2394.
(63) Nosé, S.; Klein, M. *Mol. Phys.* **1983**, *50*, 1055–1076.
(64) Hoover, W. *Phys. Rev. A: At. Mol. Opt. Phys* **1985**, *31*, 1695–1697.
(65) Wolinski, K.; Hinton, J. F.; Pulay, P. *J. Am. Chem. Soc.* **1990**, *112*, 8251–8260.
(66) Zhao, Y.; Truhlar, D. G. *Theor. Chem. Acc.* **2007**, *120*, 215–241.
(67) Yu, W.; He, X.; Vanommeslaeghe, K.; MacKerell, A. D. *J. Comput. Chem.* **2012**, *33*, 2451–2468.

Supporting information for: Towards a Broadly Applicable Force Field for d^6-piano stool complexes

Maurus Schmid,[†,‡] Thomas R. Ward,[*,†] and Markus Meuwly[*,‡]

Department of Chemistry, University of Basel, Spitalstrasse 51, CH-4056 Basel, Switzerland, and

Department of Chemistry, University of Basel, Klingelbergstrasse 80, CH-4056 Basel, Switzerland

E-mail: thomas.ward@unibas.ch; m.meuwly@unibas.ch

Contents

---

[*]To whom correspondence should be addressed
[†]Department of Chemistry, University of Basel
[‡]Department of Chemistry, University of Basel

In the following, the quality of the potential energy surface (PES) fit is discussed for selected coordinates. Energy conservation is demonstrated and the results are compared to calculations using CGenFF. Finally, all parameters and topologies required for simulations with the VB_{sp^3}, VB_{LC}, $C-VB_{LC}$ or conventional CHARMM force fields are summarized. "Assigned" parameters have equilibrium geometries from X-ray structures and force constants are set to 50 kcal/(mol∗rad^2). First is a section with the supporting Figures, then all the CHARMM force field parameters, then follows a section with all the topologies and a sample input file for complex **4**. The non-standard VBT parameters are collected in the main manuscript.

PES Fits

Figure 1 presents the quality of the fit for the bonded parameters for bonds and angles involving iridium. As described in the main text, the parameters were fitted to reproduce total energies of the complex (and not fits to individual deformations) while scanning the individual coordinates.
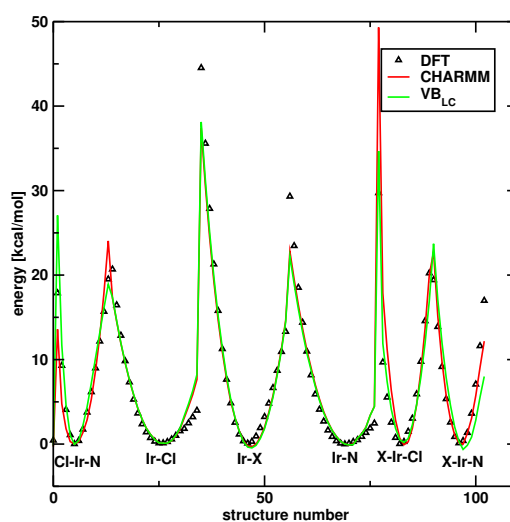
Figure 1: Quality of the PES fit for the CHARMM and VB_{LC} energy surface (lines) compared to reference DFT energies (symbols). Labels correspond to the coordinate scanned; the $x-$axis is the index of the respective structure.

S2

Figure 2 compares the reference DFT energies with the individual force field terms using the fitted parameters. As the fitting was carried out for total energies and not for individual force field terms, the fits represent the reference energies less accurately. This is because additional force field terms (including van der Waals and electrostatics) contribute to the total energy function.

(a) Ir-Cl

(b) Ir-X

(c) Ir-N

(d) X-Ir-Cl

(e) N-Ir-Cl
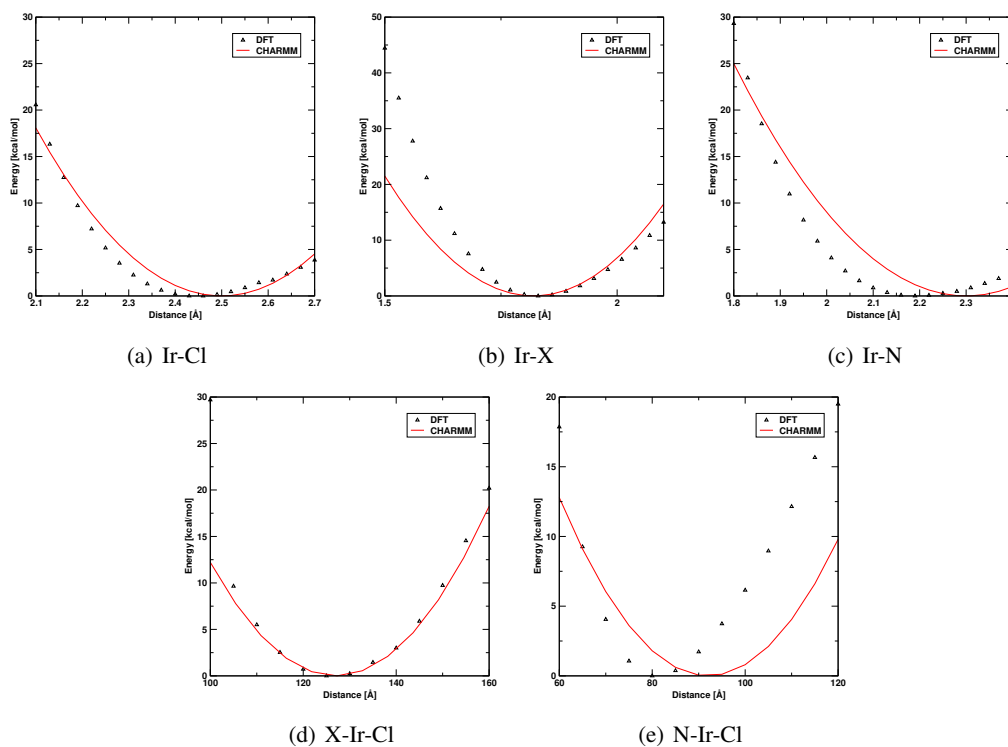
Figure 2: Comparison of the reference DFT energies with force field terms and using the fitted parameters for bonded terms involving Ir.

S3

Scanned Angles

The list of scanned metal angles (L-M-L and X-M-C$_{\text{arene-cap}}$) with the resulting CHARMM parameters is compiled in Table 1.

Table 1: L-M-L and X-M-C$_{\text{arene-cap}}$ angles scanned in the PES.

Atoms			Angle Parameters		Fitting
			Force constant kcal/mol/rad^2	Equilibrium °	on structure
N	Ir	Cl	41	92	**1**
N	Ir	X	23	115	**1**
X	Ir	Cl	55	127	**1**
N	Ir	N	45	94	**1**
N	Rh	Cl	49	100	**3**
N	Rh	X	10	110	**3**
X	Rh	Cl	55	125	**3**
N	Rh	N	72	97	**3**
N	Ru	Cl	58	94	**2**
N	Ru	X	48	134	**2**
X	Ru	Cl	20	124	**2**
N	Ru	N	42	94	**2**
Cl	Ru	P	12	59	**7**
Ir	X	C	30	90	**1**
Ru	X	C	40	90	**2**

S4

Metal Distances

The metal-ligand distances were calculated and compared between the reference X-ray and the computed structures. Average differences range from 0.05 to 0.1 Å. The maximum difference is 0.26 Å for the Ru-X bond in complex **7** using the simple VB_{sp^3} model. Table Table 2 compares the different models and Figure Figure 3 provides a graphical representation of the bond-length distribution functions from the MD simulations.

Table 2: The average and maximum differences between the experimental (X-ray) and calculated metal-ligand distances. All values are in Å.

Structure		DFT	VB_{sp^3}	VB_{LC}	C-VB_{LC}	CHARMM
4	average	0.04	0.05	0.05	0.06	0.06
	max	0.07	0.06	0.06	0.06	0.06
5	average	0.05	0.11	0.10	0.10	0.10
	max	0.10	0.15	0.13	0.16	0.16
6	average	0.04	0.10	0.08	0.09	0.09
	max	0.10	0.17	0.13	0.13	0.12
7	average	0.04	0.16	0.13	0.13	0.12
	max	0.10	0.26	0.16	0.17	0.15
8	average	0.04	0.08	0.07	0.07	0.07
	max	0.10	0.14	0.15	0.14	0.14
9	average	0.08	0.11	0.11	0.11	0.11
	max	0.18	0.15	0.14	0.14	0.15
10	average	0.07	0.10	0.09	0.08	0.08
	max	0.16	0.18	0.17	0.16	0.16
11 [a]	average	0.07	0.05	0.03		
	max	0.20	0.10	0.05		
12	average	0.04	0.07	0.05	0.06	0.06
	max	0.06	0.10	0.08	0.08	0.08
13 [a]	average	0.05	0.12	0.12		
	max	0.10	0.20	0.20		

[a] The calculations were performed only using Valbond.

S5
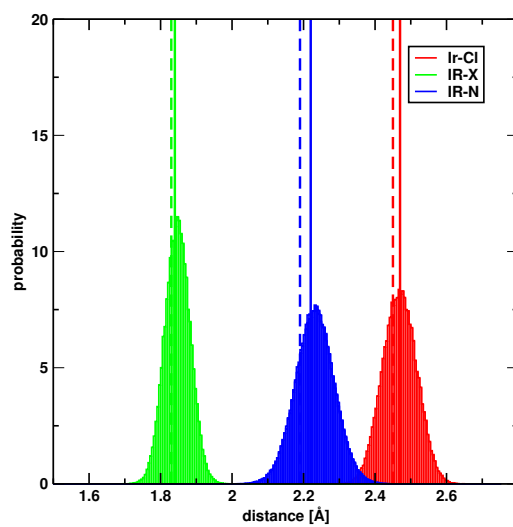
Figure 3: Metal-ligand distances for complex **1**. Equilibrium DFT (dashed line) and VB_{LC} (solid line) separations are compared with the bondlength distribution from a 2 ns gas phase simulation.

S6

Energy Conservation

The total energy is conserved during the 2 ns *NVE* VB molecular dynamics simulation and the temperature as well as the kinetic and potential energies are Gaussian distributed as expected, see Figure 4.
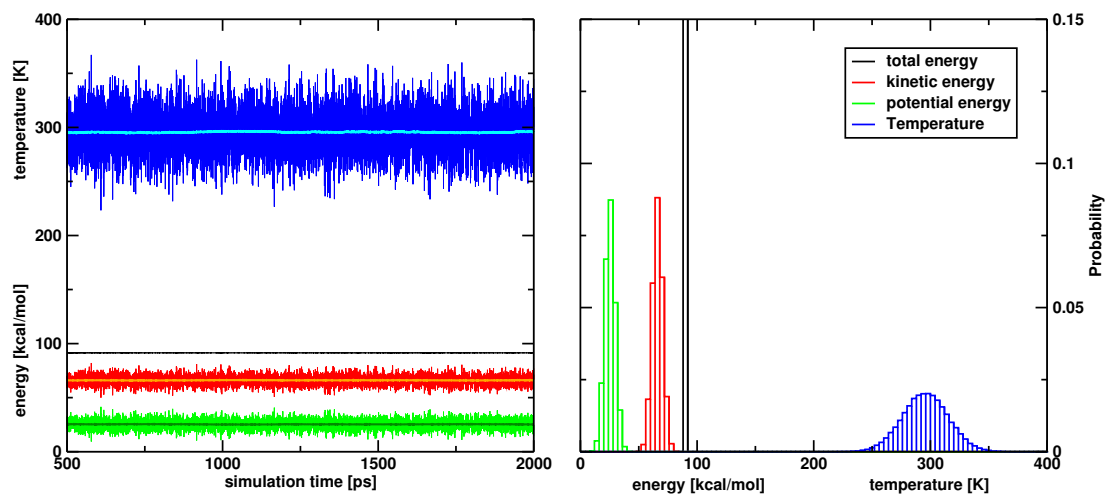
Figure 4: Computed energies for complex **4** (time course, left and histogram, right) obtained during the Valbond molecular dynamics simulation using VB_{LC}.

S7

Comparison with a CGenFF Parametrization

Table 3: Comparison of the RMSD (in Å) between X-ray and optimized structures based on the ligand parametrizations in the present work and CGenFF parameters for complexes **6** and **9**. Also, comparisons for different charge assignment schemes are reported.

		VB_{sp^3}			VB_{LC}		
		Mulliken	ChelpG	ESP	Mulliken	ChelpG	ESP
6	Ru	0.46	0.43	0.40	0.29	0.26	0.25
CGenFF **6**		1.17^b	0.47	0.44	0.37	0.35	0.31
9	Ir	0.23	0.38	0.39	0.15	0.24	0.23
CGenFF **9**		0.26	0.38	0.37	0.19	0.21	0.20

[a] The calculations were performed only using Valbond. [b] The arene cap is rotated.

S8

Interaction energies with water

Table 4: Comparison of the RMSD (in Å) between optimized (DFT and VB$_{LC}$ model with different point charges) structures of complex **1** with three, four and five nearby water molecules. The M06-2X structure is taken as the reference.

# waters	M06-2X	B3LYP	Mulliken	ChelpG	ESP
3	ref	0.09	0.41	0.28	0.28
4	ref	0.48	0.32	0.41	0.41
5	ref	0.40	0.32	0.54	0.53

The correlation between the DFT and FF (VB$_{LC}$) energies for snapshots from the MD with one or three waters (Figure 5).

(a) one water   (b) three water
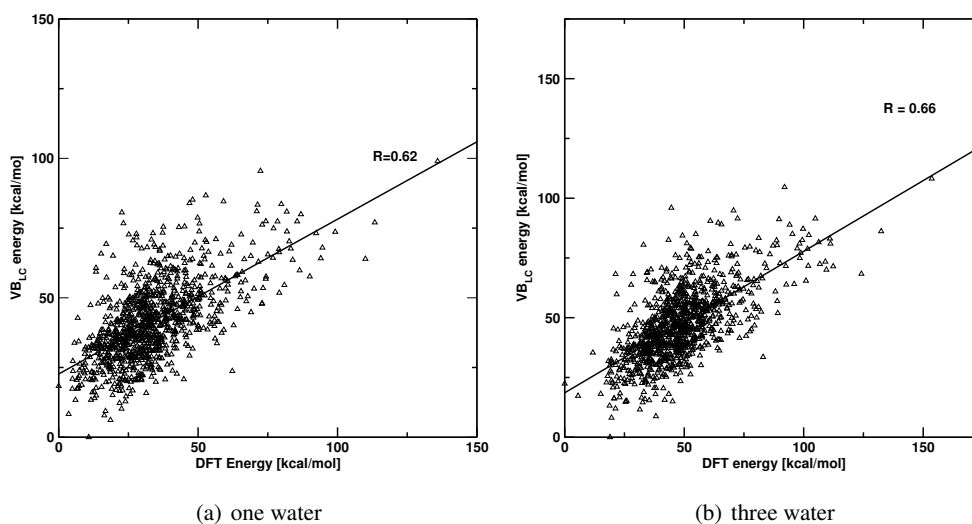
Figure 5: Correlation between DFT and VB$_{LC}$ energies for complex **1** with one (a) or three (b) nearby water molecules.

S9
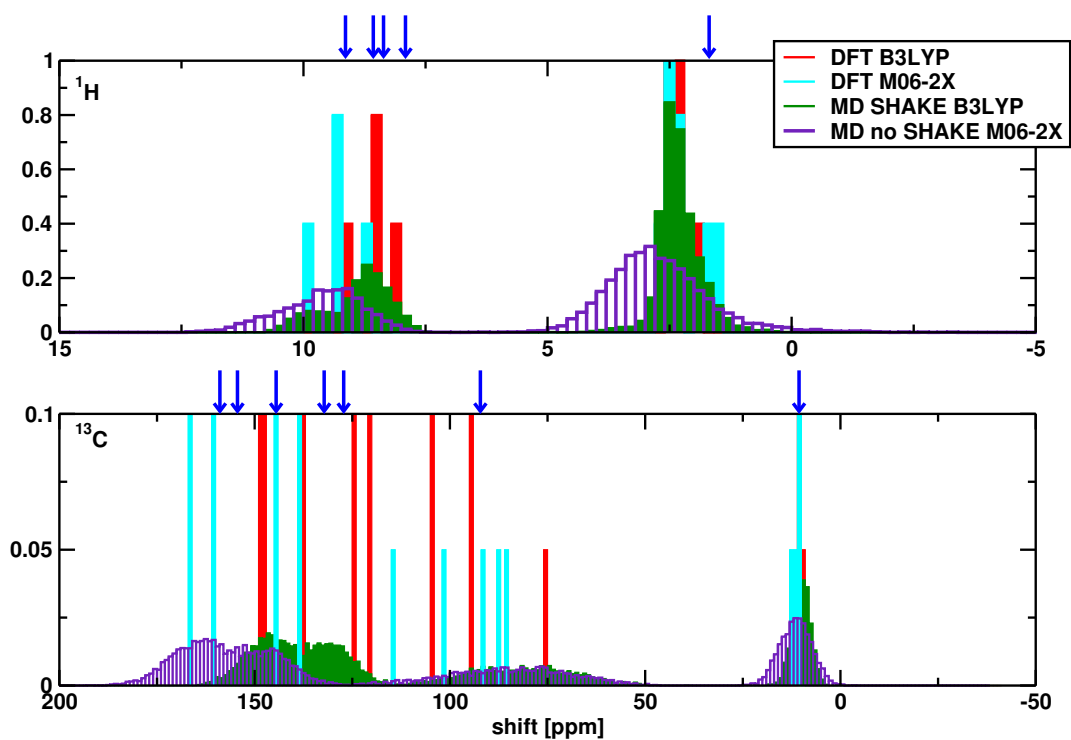
Figure 6: Calculated vs. experimental[1] (blue arrows) ^{1}H and ^{13}C chemical shifts for complex **9**. [13]C shifts were shifted relative to the lowest peak. Shifts are calculated for DFT-optimized structures (B3LYP and M06-2X functionals) and from an ensemble of MD structures with (full bars) and without SHAKE (empty bars) using both DFT methods.

7. Simulation of Transition Metal Complexes in Proteins

As all the necessary parts for a simulation of artificial metaloenzymes are in place, they are combined to yield the goal of simulating full protein-metal-ligand systems. The application of the Valbond force field on a large scale protein system is demonstrated and it is shown what informations can be gained from such simulations.

7.1. Molecular Dynamics Simulations of d^6 Transition Metal Complexes in a Protein Environment

The simulations in chapter 5, although successful, pointed out the need of a well-parametrized force-field as described in chapter 6. Validation of the more elaborate approach was performed on systems involving three different proteins and five different metal complexes.

Of the several approaches developed in chapter 6, the VB_{LC} method was chosen for the simulations for the protein complex embedded in the protein, as this method proved reliable and easily extendable. Valbond-CHARMM force field was only capable to run on a single CPU, changes to the source code were required to make the Valbond part compatible to the parallel version of CHARMM. This was achieved by limiting Valbond calculations to one CPU. But as most of the system and most calculations (including the computationally expensive nonbonded interactions) are performed with classical

CHARMM, this does not really afford a bottle-neck. Further, an interface for convenient manual inclusion and exclusion of atoms into the Valbond section was added.

A manuscript is in preparation.[164]

Molecular Dynamics Simulations of d^6 Transition Metal Complexes in a Protein Environment

Maurus Schmid,[†,‡] Thomas R. Ward,[*,†] and Markus Meuwly[*,‡]

Department of Chemistry, University of Basel, Spitalstrasse 51, CH-4056 Basel, Switzerland, and

Department of Chemistry, University of Basel, Klingelbergstrasse 80, CH-4056 Basel, Switzerland

Introduction

There are several reasons one might want to incorporate a metal complex into a protein. The most obvious one is to crate a strong binding ligand. As coordination to a metal is a strong binding motif, high performance inhibitors can be envisaged using metals. The highly diverse structures of transition metal complexes(e.g. octahedral, quadratic planar or, where we focus in this study, piano stools) can be used to achieve inhibitor shapes, that are not accessible with the classical carbon chemistry. Thus it is possible to tailor the inhibitor to a binding pocket even more than with the traditional organic molecules. Compared to traditional inhibitors, these metal complexes are much easier to synthesize, as with traditional organic chemistry inhibitors, often complicated stereospecific synthesis is required to fit the ligand to the binding site.

Another reason for incorporating transition metals into proteins is their catalytic activity. Upon insertion into the host protein, the complex remains active, but gains a stereospecific environment. These artificial metalloenzymes can achieve high enantioselectivity even for difficult reactions or

---
[*]To whom correspondence should be addressed
[†]Department of Chemistry, University of Basel
[‡]Department of Chemistry, University of Basel

1

substrates, mimicking natural enzymes.

To complement experiments, interpret observations or even predict experimental findings, computational modelling can be a valuable tool. Nowadays, computational modelling of proteins, nucleic acids or small molecules is well-established and becoming more and more routine. Numerous force fields have been developed for this purpose. Modelling of transition metal complexes has traditionally been achieved relying on *ab initio* or quantum mechanical methods like density functional theory (DFT). With *in silico* high-throughput screening in approaches mind, the computational power required with DFT becomes prohibitive. Alternatively, QM/MM may be attractive but, for library-screening purposes, such an approach is not sufficiently efficient and scalable. Instead, force field methods would be highly desirable. As force field calculations are much faster than any of the QM methods, they do not only allow screening of large libraries, but also simulations of much larger systems involving proteins including artificial metalloenzymes or enzymes with transition metal inhibitors mentioned above.

Systems

Simulation has been done with 3 different protein-piano stool systems:

Pim-1 The human kinase Pim-1 is an example of a metal complex working as a inhibitor. It is a monomer, its structure and the location of the metal complex can be found in Figures 9(a) and 9(b) Meggers *et. al* developed several high potency inhibitors based on metal complexes. This pianostool complex consists of a central ruthenium with a *Cp* ligand, a carbon monoxide ligand and a planar two-dentate structure. It was designed to mimic the structure of the known protein kinase inhibitor staurosporine.[1] Staurosporine closely matches the shape of the ATP-binding site with its planar base and globular head group. For the metal inhibitor, the globular group is replace with the more versatile and more easy synthesizable metal moiety.

Streptavidin Streptavidin (Sav) is a protein with an extraordinarily high binding affinity to biotin.

<div align="center">2</div>

It is a tetramer, consisting of four identical monomers. All four of these monomers can bind a biotinolated ligand thus one protein can complex four ligands. The binding sites for two of these monomers face each other, whereas the two other two binding sites lie on the opposite side of the protein. Figures 7(a) and 7(b) It is used as a host for catalytically active metal complexes by our group.The metal complexes are incorporated into the host protein using a linker that connects the biotin moiety with the metal. The length of the linker has a large influence on the catalyzed reaction as it determines the position of the metal relative to the protein. Another main target for designing the artificial metalloenzymes is by mutating especially the residues in close vicinity to the metal. The streptavidin-biotin interaction is one of the most stable non covalent bonds known. This is exploited to attach the metal complex to the protein with an biotin linker. A ruthenium is the center of the piano stool. A chloride and a *para*-cymene are the remaining metal-ligands.

hCa II Human carbonic anhydrase II (hCa II) is a member of the family of carbonic anhydrases, that catalyze the reversible hydration of carbon dioxide. Isoforms are involved in various pathological processes including infections, tumorigenicity, osteoporosis, epilepsy, obesity, gluconeogenesis, lipogenesis, ureagenesis, or glaucoma. Metalcomplexes are either inhibitors for the catalytic function or are being used as catalysts similar as in streptavidin. Figures 8(a) and 8(b) Arylsulfonamides are well-known inhibitors for carbonic anhydrases. They bind to the Zinc located in the center of the binding pocket. The metal complex in here is an iridium piano stool complex. Bound to the metal are a chloride, a Cp^* group and a sulfonamide linker providing the connection to the protein. Recently we did a ligand binding study involving these ligands.[2]

All of the piano stool complexes have stereocenters at the metal and thus have at least one mirror image. Where a crystal structure was available the absolute configuration present was used in the study. For the Sav **2** models however, both enantiomers were considered and compared.

3

Methods

System Setup

Each system was partitioned into two parts, a Valbond and a CHARMM component. Whereas most of the atoms were assigned to the CHARMM region, the ligands were calculated using Valbond. Therefore the CHARMM region consisted of the protein and the solvent atoms.

Where available, X-ray structures were taken as a starting point for the calculations (Pim-1: PDB code 2BZH, Sav: PDB code 2WPU, hCa: PDB code 3PYK). For the systems where no crystal structure was available, the protein of another X-ray structure (1G54 for hCa II, 1mk5 for Sav) was taken as host protein and the ligand was docked manually to yield the typical binding mode found in the proteins. This approach has proven to yield reliable structures for biotin and hCa II[2,3] as the biotin respectively the arylsulfonamide are very strong binding motifs and are very conserved between different crystal structures.

Valbond

Valbond (VB)[4–6] is based on valence bond theory as originally developed by Pauling.[7] It replaces the conventional harmonic bending term in a general-purpose force field with the VB-energy expression which is based on hybrid orbital strength functions as the basis for a molecular mechanics expression. These functions not only describe the energy of bond angles around the minimum, but also at very large distortions where the harmonic approximation breaks down. In this fashion, VB can reproduce unusual geometries as illustrated by the distorted trigonal prism reported for $[W(CH_3)_6]$.[8] We extended the Valbond formalism to include the *trans*-influence which yields Valbond-trans (VBT)[9] and recently we parametrized Valbond for d^6 piano stool complexes.[10] For a more detailed description of the Valbond formalism, refer to the original literature[4–6] and our previous contributions[9,10]

In our implementation of VB and VBT[9] (available from CHARMM[11] version 37 onward), the conventional CHARMM angle terms are replaced by hybrid orbitals which can be assigned by

4

the user. In this way, VB and VBT can be combined and are compatible with the conventional CHARMM force field and simulations in a variety of ways are possible, as will be explained further below.

Simulations

Molecular mechanics simulations were carried out in CHARMM[11] supplemented with the VB module[9] using the CHARMM27[12] force field with custom parameters[2]. For biotin we relied on parameters developed by Schulten *et. al.*[13]

The proteins were solvated in a cubic box of explicit water (TIP3P model[14]) with 80 Å edge length. The water molecules had been equilibrated previously and periodic boundary conditions were applied. The resulting systems consisted of approximately 50000 atoms each.

Molecular dynamics simulations have been performed for systems using Valbond-CHARMM. Simulation was done in the *NVE* ensemble using a step-size of 1 fs. The recommended CHARMM27 cutoffs and settings[12] were used in the simulations: A switching (vdW terms) respectively shifting (electrostatics) function between 10 Å and 12 Å, cutoffs at 12 Å and atom based electrostatics.

Analysis

The trajectory was reoriented and recentered on the protein backbone of the crystal structure to remove translational and rotational motion. Root mean square deviations (RMSD) were calculated and density maps were generated from the trajectories by analyzing the location of the atoms during the simulation. RMSDs were done for the protein backbone, protein heavy atoms (no hydrogens) and for the ligands separately. Whereas the RMSD for the backbone describe large scale movement, the RMSD for the protein heavy atoms also includes side chains movements. For the ligands only heavy atoms were considered. The density maps were calculated by taking snapshots every picosecond which resulted in a total of 10000 structures for a trajectory of 10 ns. The resulting density maps gave a good indication of the movement of the atoms during the simulation, highlighting more

5

or less mobile areas not only of the protein, but also of the ligand. Comparison between the apo- and the holoprotein could give insight into the influence of the cofactor and maybe point out some allosteric motions.

Dynamic Cross Correlation Maps (DCCM) are a visual representation for the movement of the residues relative to each other. The fluctuation of the residues during the molecular dynamics simulation are calculated for each residue and correlated to each other. A high correlation is the result of concerted movement caused by related residues, low correlation indicates independent movement, typical for residues far away from each other. Negative correlation is possible and is originating in opposing movement. Then the fluctuations of the residues were cross-correlated to each other for each window. A typical "fingerprint" for each protein emerges, as the residues correlating with each other are mainly given by close contacts, which in turn are a result of the primary, secondary, tertiary or quarternary structure.

Binding Free Energy The binding free energy was calculated using the Molecular Mechanics Generalized Born Surface Area (MM-GBSA[15]) approach. In MM-GBSA the ligand binding free energy ΔG is decomposed in the following way:

$$\Delta G_{\mathrm{bind}} = \Delta E_{\mathrm{MM}} + \Delta G_{\mathrm{solv}} - T\Delta S_{\mathrm{MM}} \tag{1}$$

Eq. 1 describes the computation of a binding free energy according to a thermodynamic cycle which includes the enthalpic (E_{MM}) contribution for protein-ligand interaction in the gas phase and the desolvation free energies for the separated and combined protein-ligand complex, respectively. E_{MM} is the "gas phase" energy which is calculated with the CHARMM27[16] force field. E_{MM} contains all internal, electrostatic, and van der Waals energies and the nonbonded interactions are computed without cutoff.

The solvation part ΔG_{solv} consists of a solvent-solvent cavity term (G_{cav}), a solute-solvent van

6

der Waals term (G_{vdW}) and a solute-solvent electrostatic polarization term (G_{pol}), i.e.

$$G_{solv} = G_{cav} + G_{vdW} + G_{pol} \tag{2}$$

The polarization term in MM-GBSA is calculated within the generalized Born approximation

$$G_{pol} = -166 \left(1 - \frac{1}{\varepsilon}\right) \sum \sum \frac{q_i q_j}{f_{GB}} \tag{3}$$

where ε is the dielectric constant of water ($\varepsilon = 80$), q_i and q_j are the charges of atoms i and j. f_{GB} is an expression that depends on the Born atom radii α_i and α_i and distances r_{ij}.[17]

$$f_{GB} = r_{ij}^2 + \alpha_i \alpha_j \exp\left(\frac{-r_{ij}^2}{8\alpha_i \alpha_j}\right) \tag{4}$$

G_{vdW} and G_{cav} are assumed to be linearly related to the solvent accessible surface area (SASA):

$$G_{cav} + G_{vdW} = \sum \sigma_k SA_k \tag{5}$$

From the MD simulations snapshots were recorded every 10 ps and the above contributions to ΔG were computed. Then, the energies of these snapshots were averaged to yield final energies and their fluctuations.

$\Delta E_{MM} + \Delta G_{solv}$ can be decomposed into per-residue contributions ΔG_i to the overall $\Delta G = \sum_i^N \Delta G_i$. Such a decomposition allows to trace the changes between simulations back to individual residues. This is important in order to identify suitable mutations to enhance or decrease ligand-binding affinities.

It has been previously shown that the influence of S_{trans}, S_{rot} and S_{vib} on *relative* binding free energies $\Delta\Delta G$ is in general small when considering only similar ligands[18–22] or may be even detrimental.[2] Further, as we were mainly interested in residue contribution and a per residue decomposition of S_{trans}, S_{rot} and S_{vib} is not meaningful, we omitted calculation of further entropic

7

terms.

All individual energy contributions ε_i are calculated according to the thermodynamic cycle[19] separately for the entire system, the protein and the ligand. Then the difference $\Delta\varepsilon_i$ between the protein-ligand complex and the separated system is computed for every component:

$$\Delta\varepsilon_i = \varepsilon_i^{\text{Prot}-\text{Lig}} - (\varepsilon_i^{\text{Prot}} + \varepsilon_i^{\text{Lig}}) \tag{6}$$

Results

Structural effects

The RMSD can give a measure for the stability of the simulated system. The higher the RMSD, the more the structure drifted away from the starting structure.

During the 10 ns simulations, all simulated systems remained relatively close to the initial structure. The average RMSD for the backbone remained between 0.9 Å for hCa2 and 2.1 Å for Pim-1 without ligand. Simulations lacking any ligand were less stable. This might be a result of the choice for the initial structures as all of the proteins were crystallized containing a ligand. For Pim-1 this effect was most pronounced. Without the presence of the piano stool complex, the binding site was opening up. The process was not yet finished at the 10 ns mark as the RMSD was still increasing.

The RMSD including the sidechains were only slightly higher than for the backbone (0.3 Å to 0.5 Å).

The most stable metal complex was the one for Pim-1. It has a very rigid structure which did not allow for a lot of movement. The ligands in each of the four monomers of Sav did not behave identically. For the Sav **1** system a range of 0.7 Å was observed. The docked structures of Sav **2a** and Sav **2b** were on average less conserved than the simulations with a crystal structure.

However, it is hard to compare the RMSD of the ligands between the different proteins as the are quite different.

8

The density map surfaces highlight areas of high and low stability during the simulation. When compared to the solvent accessible surfaces (sas) of the crystal stuctures, gaps are much more filled. Where the atoms were moving a lot during the simulations, cone-like structures could be identified. The proteins were more conserved in the center than in the periphery. Some cofactors were quite labile, whereas **Sav 1** seemed to be more conserved than **Sav 2**, especially in **Sav 2b** the whole vestibule tended to be filled and close contact between two adjacent cofactors was made. The large movement in **Sav 2b** indicated a lower stability than for **Sav 2a** which was supported by the MM-GBSA calculations (see below.)

Dynamic Cross Correlation Maps

In the DCCM mainly residue-residue interactions can be observed. Correlations below 0.4 are considered insignificant and were not considered here as they are not reliably reproducible. On the diagonal line are self correlations, i. e. correlation of a residue with itself, which are always 1. Left of it is correlated (positive) movement, on the right is anticorrelated (negative) movement. Anticorrelated movement was only observed for the **Pim nl** simulation.

As the cofactors typically interact with several residues, no high correlation to a single residue had to be expected but rather correlation to several residues. But the residue-residue correlation seems to dominate the residue-ligand correlation and the metal complexes correlate only weakly with the residue movement, only for **Sav 2b** and **Pim-1 1** weak correlation is visible.

Binding Free Energy

References

(1) Debreczeni, J. E.; Bullock, A. N.; Atilla, G. E.; Williams, D. S.; Bregman, H.; Knapp, S.; Meggers, E. *Angew. Chem. Int. Ed.* **2006**, *45*, 1580–1585.

(2) Schmid, M.; Nogueira, E. S.; Monnard, F. W.; Ward, T. R.; Meuwly, M. *Chem. Sci.* **2012**, *3*, 690.

9

(3) Zimbron, J. M.; Heinisch, T.; Schmid, M.; Hamels, D.; Nogueira, E. S.; Schirmer, T.; Ward., T. R. *submitted to JACS* **2012**,

(4) Root, D. M.; Landis, C. R.; Cleveland, T. *J. Am. Chem. Soc.* **1993**, *115*, 4201–4209.

(5) Cleveland, T.; Landis, C. R. *J. Am. Chem. Soc.* **1996**, *118*, 6020–6030.

(6) Landis, C. R.; Cleveland, T.; Firman, T. K. *J. Am. Chem. Soc.* **1998**, *120*, 2641–2649.

(7) Pauling, L. *J. Am. Chem. Soc.* **1931**, *53*, 1367–1400.

(8) Landis, C. R.; Cleveland, T.; Firman, T. K. *J. Am. Chem. Soc.* **1995**, *117*, 1859–1860.

(9) Tubert-Brohman, I.; Schmid, M.; Meuwly, M. *J. Chem. Theory Comput.* **2009**, *5*, 530–539.

(10) Schmid, M.; Ward, T. R.; Meuwly, M. *submitted to JCTC* **2012**,

(11) Brooks, B. R.; Bruccoleri, R. E.; Olafson, D. J.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187–217.

(12) MacKerell, A. D. et al. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.

(13) Izrailev, S.; Stepaniants, S.; Balsera, M.; Oono, Y.; Schulten, K. *Biophys. J.* **1997**, *72*, 1568–1581.

(14) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–936.

(15) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. *J. Am. Chem. Soc.* **1990**, *112*, 6127–6129.

(16) MacKerell Jr., A. D.; Brooks III, C. L.; Nilsson, L.; Roux, B.; Won, Y.; Karplus, M. In *CHARMM: The Energy Function and Its Parameterization with an Overview of the Program*; v. R. Schleyer et al., P., Ed.; The Encyclopedia of Computational Chemistry; John Wiley & Sons: Chichester, 1998; Vol. 1; pp 271–277.

10

(17) Lee, M. S.; Feig, M.; Salsbury, F. R.; Brooks, C. L. *J. Comput. Chem.* **2003**, *24*, 1348–1356.

(18) Wang, W.; Kollman, P. a. *J. Mol. Biol.* **2000**, *303*, 567–582.

(19) Gohlke, H.; Kiel, C.; Case, D. A. *J. Mol. Biol.* **2003**, *330*, 891–913.

(20) Zoete, V.; Meuwly, M.; Karplus, M. *Proteins* **2005**, *61*, 79–93.

(21) Thorsteinsdottir, H. B.; Schwede, T.; Zoete, V.; Meuwly, M. *Proteins: Struct., Funct., Bioinf.* **2006**, *65*, 407–423.

(22) Hou, T.; Wang, J.; Li, Y.; Wang, W. *J. Chem. Inf. Model.* **2011**, *51*, 69–82.

11

Tables

Figures

Probability Distribution of Atoms During Simulation

12

Table 1: Summary of the simulations. RMSD bb includes only backbone atoms, RMSD and RMSD ligand exclude hydrogens.

Protein	System	PDB code	monomer	RMSD bb Å	RMSD Å	RMSD ligand Å	$\Delta E_{\mathrm{MM}} + \Delta G_{\mathrm{solv}}$ kcal/mol	σ kcal/mol
Sav	Sav nl	1MK5		1.23	1.55	-	-	11.94
			Sav1		1.52	-	-	5.88
			Sav2		1.50	-	-	6.90
			Sav3		1.66	-	-	6.74
			Sav4		1.47	-	-	6.31
	Sav 1	2WPU		1.17	1.51	1.99	-181.28	14.07
			Sav1		1.49	1.54	-47.95	6.81
			Sav2		1.57	2.04	-39.77	6.53
			Sav3		1.53	1.86	-48.46	6.89
			Sav4		1.45	2.26	-45.78	7.98
	Sav 2a	1MK5		1.18	1.48	2.27	-195.26	14.09
			Sav1		1.58	1.99	-47.27	6.58
			Sav2		1.31	1.86	-56.14	7.02
			Sav3		1.62	1.68	-55.39	6.31
			Sav4		1.40	3.15	-42.52	6.21
	Sav 2b	1MK5		1.22	1.53	2.77	-187.20	
			Sav1		1.57	2.15	-51.17	
			Sav2		1.53	2.24	-46.34	
			Sav3		1.50	3.86	-46.34	
			Sav4		1.48	2.14	-47.37	
hCa II	hCa 2 nolig	1G54		1.00	1.47	-	-	
	HCa 1	3PYK		0.98	1.45	1.69	-29.95	5.76
PIM1	PIM1 nolig	2BZH		2.13	2.48	-	-	
	PIM1	2BZH		1.37	1.81	1.27	-28.87	5.53

(a) **Sav-nl**

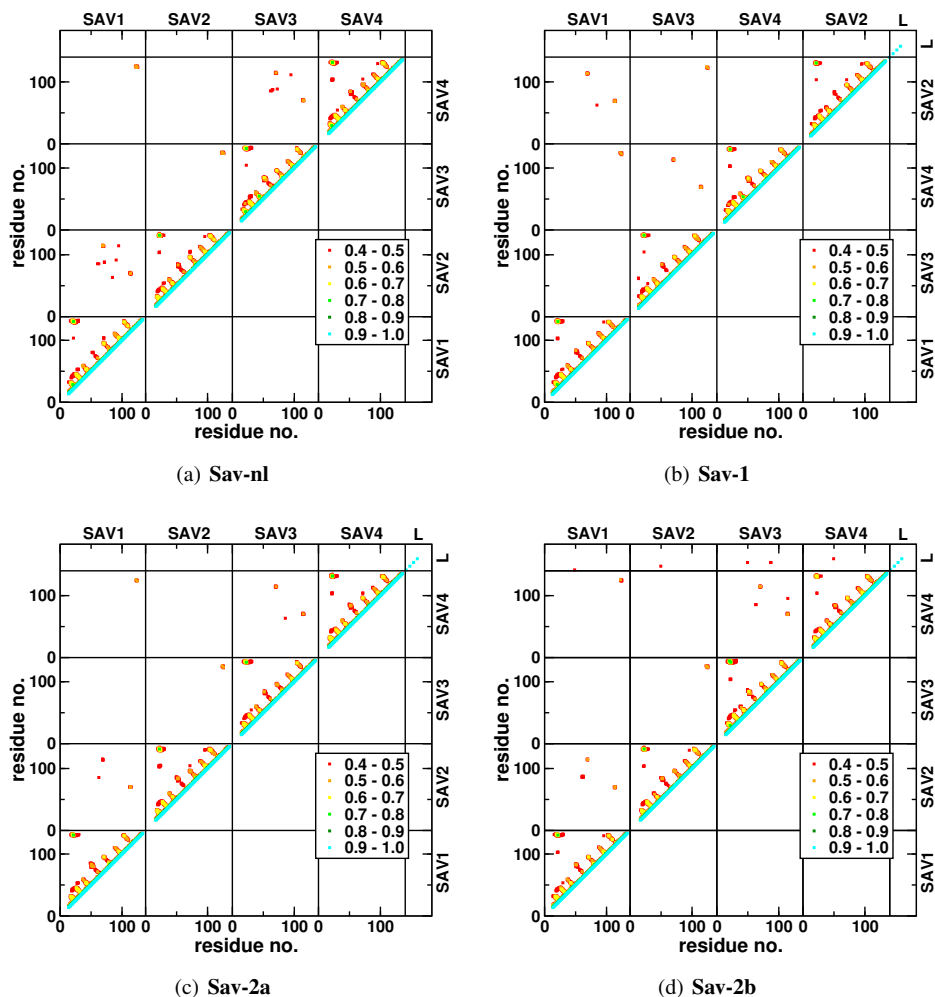(b) **Sav-1**

(c) **Sav-2a**

(d) **Sav-2b**

Figure 1: Dynamic Cross Correlation Map for Sav. On the diagonal line are the self correlations which are always 1, left of it is correlated (i.e. positive) movement, on the right would be anticorrelated (i.e. negative) movement, but none was observed. The Sav monomers are separated by black bars. Correlations below 0.4 are considered insignificant and are not displayed here. Except for Sav 2b no significant correlation between the ligand and the protein can be observed.
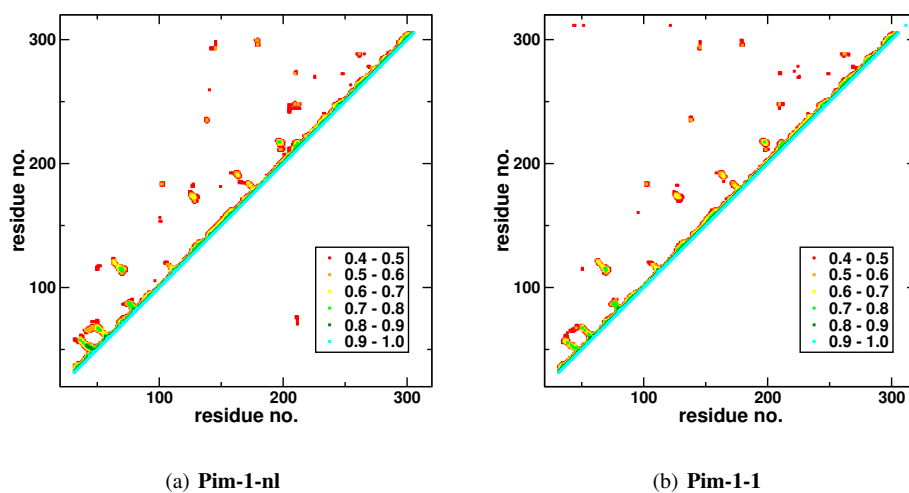
14

(a) **Pim-1-nl**

(b) **Pim-1-1**

Figure 2: Dynamic Cross Correlation Map for Pim-1.

(a) **hCa-nl**

(b) **hCa-1**
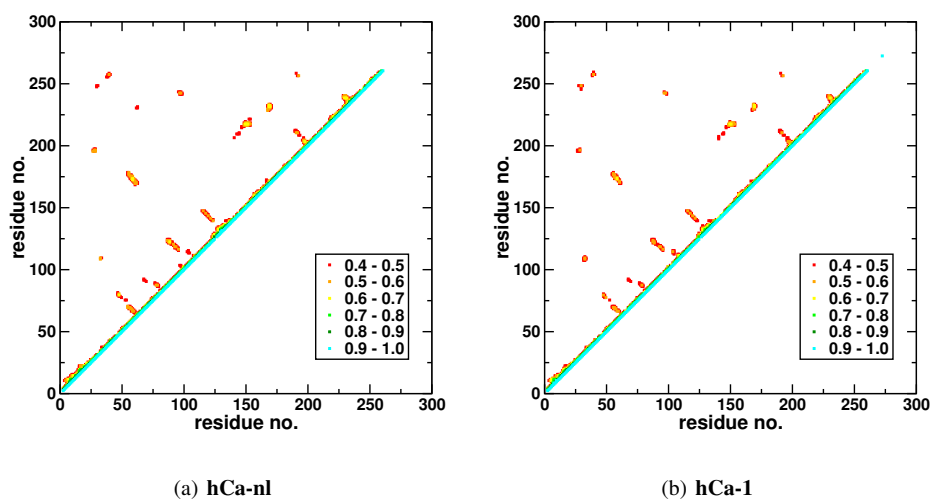
Figure 3: Dynamic Cross Correlation Map for hCa 2.

15

143

(a) **Sav-nl** sas   (b) **Sav-nl** density map   (c) **Sav-1** sas   (d) **Sav-1** density map

(e) **Sav-2a** sas   (f) **Sav-2a** density map   (g) **Sav-2b** sas   (h) **Sav-2b** density map
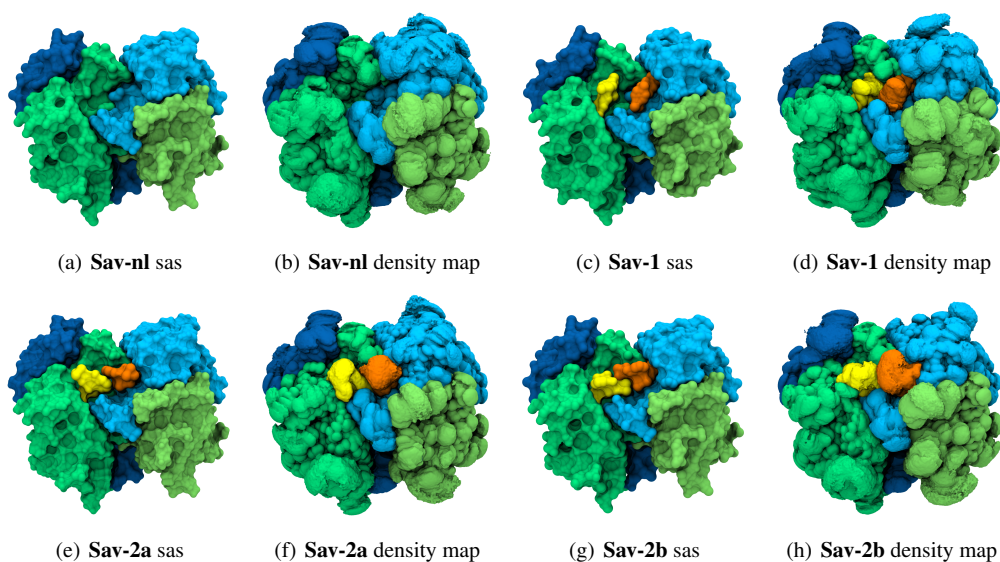
Figure 4: Comparison of solvent accessible surface area (sas) with the density map surface for the simulations involving Sav. The monomers of the protein are in shades of green and blue, the two visible cofactors are yellow or orange.

(a) **hCa-nl** sas   (b) **hCa-nl** density map   (c) **hCa** sas   (d) **hCa** density map

Figure 5: Comparison of solvent accessible surface area (sas) with the density map surface for the simulations involving hCa. The protein is in blue, the cofactor in yellow.

16

(a) **Pim-1-nl** sas   (b) **Pim-1-nl** density map   (c) **Pim-1-1** sas   (d) **Pim-1-1** density map

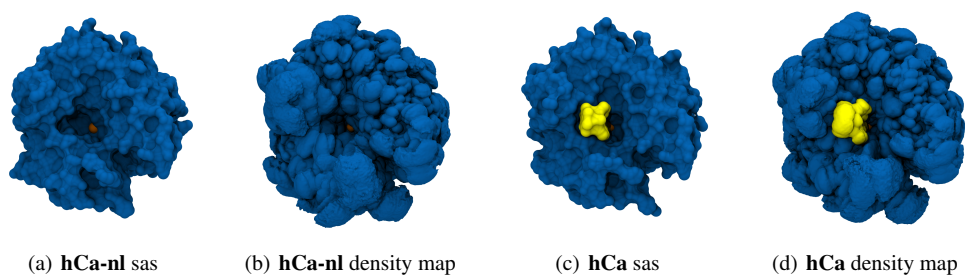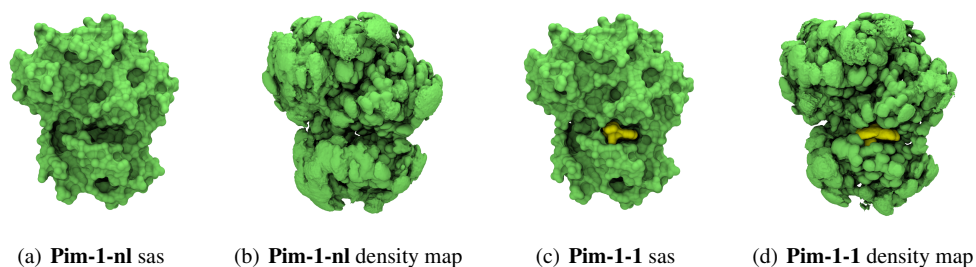Figure 6: Comparison of solvent accessible surface area (sas) with the density map surface for the simulations involving Sav. The protein is in green, the cofactor in yellow.

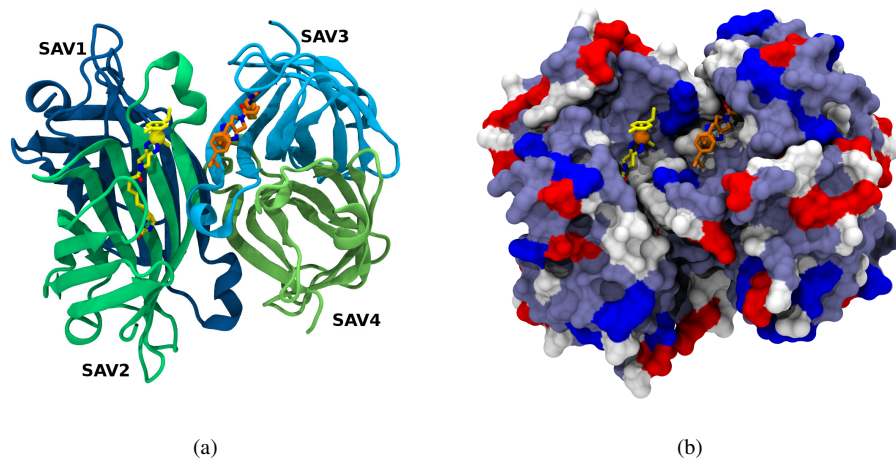(a)   (b)

Figure 7: (a): Cartoon representation of streptavidin with ligand **KYT** (PDB code 2wpu). The monomers are coloured with shades of blue and green, the C atoms of the ligands are coloured in yellow and orange respectively. (b): Solvent accessible surface representation, the surface is coloured by residue type; nonpolar white, polar light blue basic dark blue and acidic red.

17

(a)   (b)
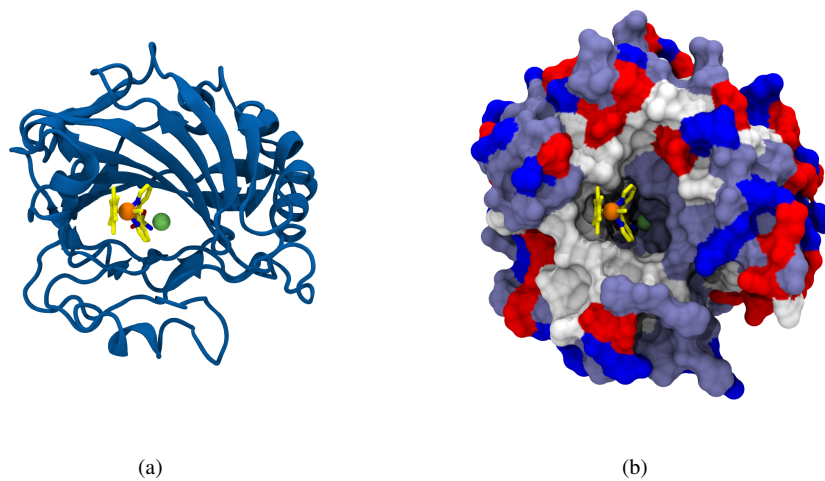
Figure 8: Cartoon (a) respectively surface (b) representation of hCa II with ligand.

(a)   (b)
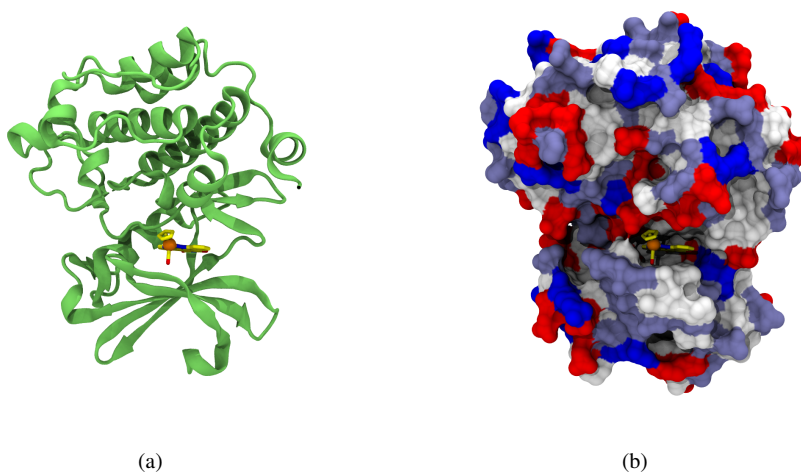
Figure 9: Cartoon (a) respectively surface (b) representation of Pim-1 with ligand.

18
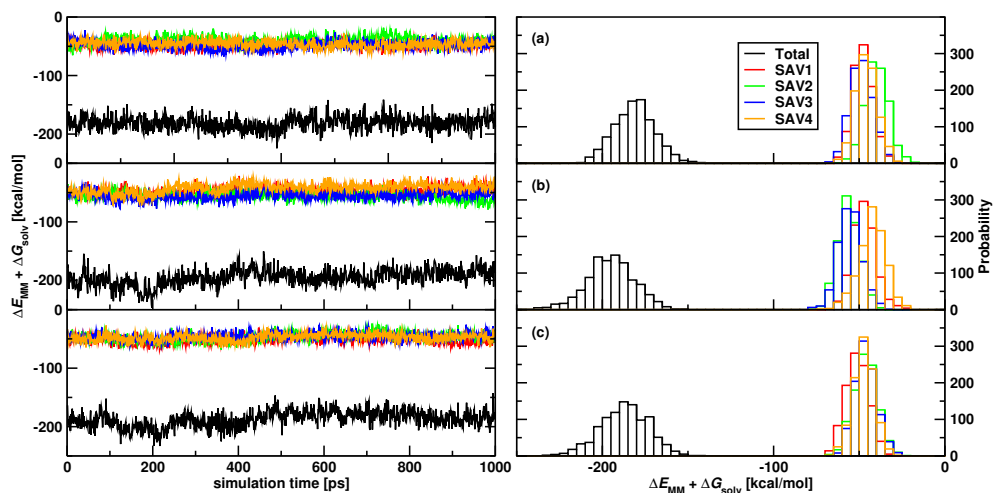
Figure 10: $\Delta E_{\mathrm{MM}} + \Delta G_{\mathrm{solv}}$ from the MM-GBSA calculation for the Sav systems.
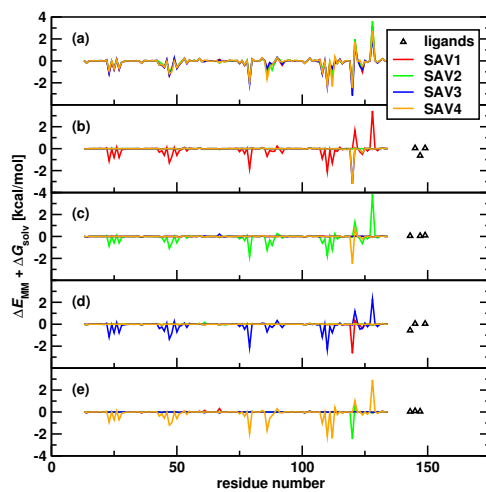
Figure 11: By-residue decomposition of $\Delta E_{\mathrm{MM}} + \Delta G_{\mathrm{solv}}$ for the MM-GBSA of the **Sav 1** system.

19
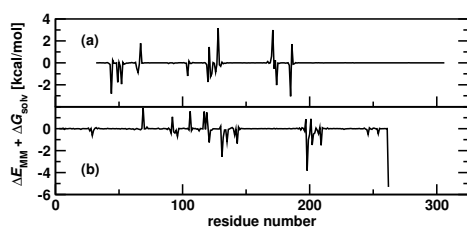
Figure 12: By-residue decomposition of $\Delta E_{\mathrm{MM}} + \Delta G_{\mathrm{solv}}$ for the MM-GBSA of the **hCa 2** and the **Pim-1** systems.

20

(a) Sav-1

(b) Sav-2a

(c) Sav-2b
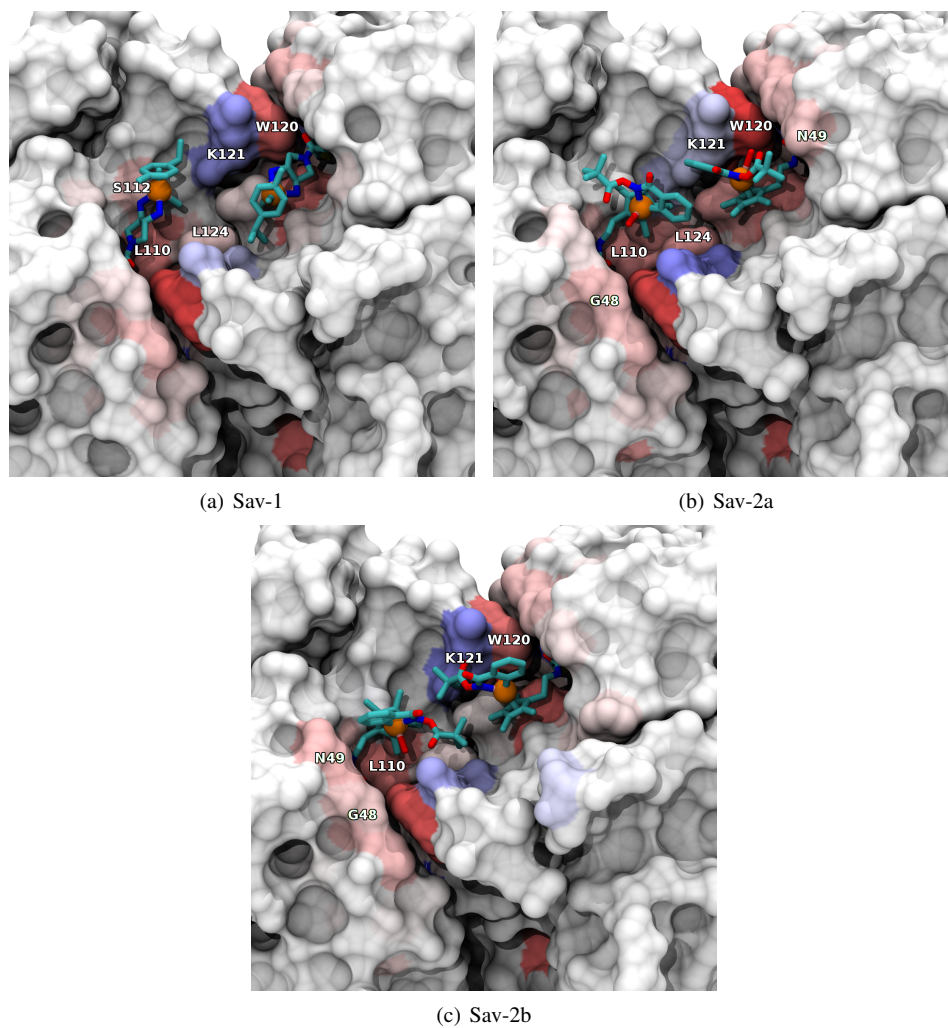
Figure 13: Binding free energy for for Sav. The SAS is coloured by the interaction energy ranging from -4 kcal/mol (red) to +4 kcal/mol (blue). Relevant residues are labelled for one of the monomers.

21

(a) Pim-1   (b) hCa 2
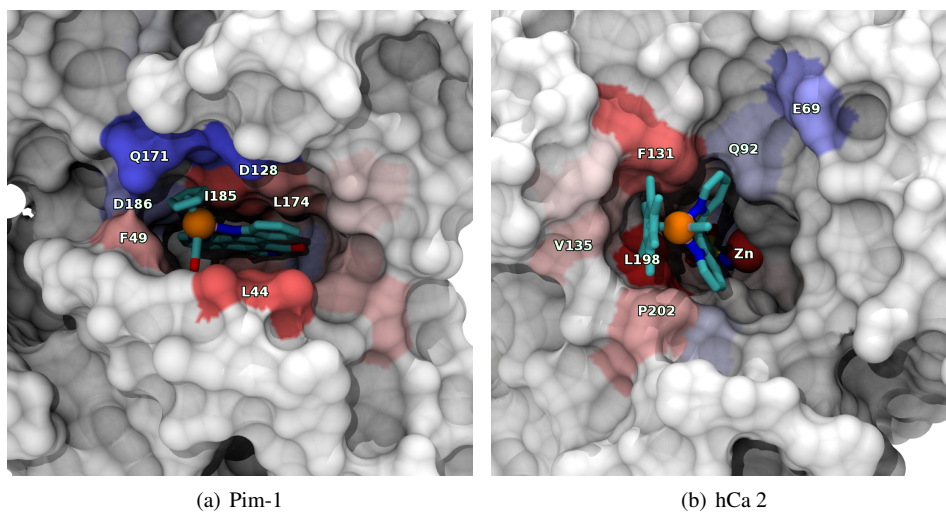
Figure 14: MM-GBSA Pim-1 and hCa 2.

22

Part III.

Conclusion and Outlook

8. Conclusion and Outlook

Artificial metalloenzymes have the potential to be applied for many different purposes. As a hybrid between a transition metal catalyst and a protein, they can inherit some properties of both. Computational tools can be invaluable for the design and understanding of these hybrid catalysts. While protein-ligand interactions have been in the focus of many computational studies, metal and especially transition metal systems like those in artificial metalloenzymes had been largely ignored and are challenging to model. Within this thesis, various computational approaches were pursued to model metal mediated protein ligand interactions. Four challenges were identified and addressed.

The first one was the identification of suitable protein scaffolds for the creation of artificial facial-triad motifs. Among natural metalloenzymes, the facial-two-histidines one-carboxylate binding motif (FTM) is a widely represented first coordination sphere motif present in the active site of a variety of metalloenzymes. Statistically representative FTMs were selected and used as template for the identification of structurally characterized proteins bearing these three amino acids in a promising environment for binding to a transition metal. While most hits consisted of either apo-FTMs or bore strong sequence homology to known FTMs, seven such structures lying within a cavity were identified as novel and viable scaffolds for the creation of artificial metalloenzymes bearing an FTM. This approach could be extended to other metal bearing motifs, as the general strategy is not limited to these motifs.

153

8. Conclusion and Outlook

Towards the simulation of artificial metalloenzymes, simulations of protein-ligand systems were performed and the second challenge was the development of reliable force field parameters for the anchoring of sulfonamide bearing ligands within carbonic anhydrase. Arylsulfonamide derivatives are widely studied high affinity inhibitors of the zinc isozyme human carbonic anhydrase II (hCA II). From molecular dynamics simulations and MM-GBSA calculations, relative binding free energies were determined for 17 previously experimentally characterized protein-ligand complexes. Decomposition of these energies led to the identification of critical amino acid residues with a significant contribution to the affinity towards the ligands. The computational vs. experimental energies were in good agreement thus suggesting that the force field parameters are useful for the *in silico* design of a wider range of carbonic anhydrase inhibitors and can be used for simulations of artificial metalloenzymes.

Computational studies towards the design of an artificial metalloenzyme were performed by docking a biotinylated three-legged rhodium piano stool complex to streptavidin to afford an artificial transfer hydrogenase. It was proposed, that dual anchoring of the metal would reduce the flexibility of the metal complex bound to the protein and thus increase the selectivity of the hybrid catalyst. The computational design towards the identification of a spacer length and the site of mutation was confirmed by X-ray crystallography. It was then shown that coordination of the catalyst precursor to a suitably positioned histidine residue has a significant impact on the catalyst's performance, both in terms of activity and of selectivity. As the crude and simple simulation employed in the dual anchoring study helped developing and understanding the system, a more elaborate and validated approach could be even more useful.

Towards this goal, more reliable parameters for metal complexes were derived. In order to predict the geometry of catalytically relevant piano stool artificial cofactors within host proteins, the third project aimed at developing force-field parameters for d^6-piano

154

stool complexes. CHARMM and the Valbond force field parameters for the calculation of these complexes have been developed and evaluated. Structure optimizations with these parameters lead to a good agreement between the the calculated force field structures and the X-ray structures, they were comparable to RMSDs obtained between X-ray and DFT optimized structures. Molecular dynamics calculations remained stable and did not drift from the observed X-ray structures. Some further work on the refinement of the parameters may be performed, especially the nonbonded interactions may not be parametrized extensively enough. The force field could be improved by fitting of the charges or the Lennard Jones parameters, or the simple point charge electrostatics with Coulomb interactions may even be replaced with a charge system that covers effects like polarizability or inhomogeneous charge distribution. Further development could be done on the repulsion dispersion terms as the 6/12 Lennard Jones term is not the best choice for these kind of interactions. Finally, through the use of non-hypervalent CHARMM, the *trans* influence is not accounted for anymore and a model for the implementation of the Jahn Teller effect needs to be developed. Nevertheless, we demonstrated that the Valbond approach can be successfully employed on complexes with ligands of the η^n with $n = 5, 6$ type and that the present force field and implementation yield good result. This approach may easily be extended to other η^n type ligands with e.g. $n = 2$ or $n = 3$.

With the foundations laid out, simulations on the full protein-metal-complex systems were possible. And finally, the developments above were combined towards the reliable prediction of first and second coordination sphere environments around artificial cofactors/inhibitors. As Valbond is compatible with the traditional CHARMM angular parameters, mixed Valbond-CHARMM calculations could be performed to combine the strength of both approaches: the established quality of CHARMM for simulations of large biomolecules and the versatility of Valbond for unusual geometries. Simulations on three proteins (streptavidin, hCa II and the human kinase Pim1) with various piano

155

stool complexes have been carried out to test the validity of this combined approach.

It has been shown that the general approach and the methods used and developed in this thesis are valid and valuable. Simulations involving metals and proteins can be performed and give valuable information for experiments. For routine application in the simulation of metal mediated protein-ligand interactions, the work-flow could be adjusted and simplifications added to be more accessible towards practical applications. As reliable calculations need a considerable amount of time, the implementation of the new parallel approach in CHARMM (DOMDEC[165,166]) or the use of the newly developed GPU based nonbonded calculation might be very beneficial and make calculations significantly faster. However, it remains to be tested if and how Valbond needs to be adapted to be compatible.

Bibliography

[1] J. Bartoll, 9th International Conference on NDT of Art, 2008, pp. 1–9.

[2] *Nobel Lectures, Chemistry 1901-1921*, Elsevier Publishing Company, Amsterdam, 1966.

[3] C. Housecroft and A. G. Sharpe, *Inorganic Chemistry (2nd Edition)*, Prentice Hall, 2004, p. 992.

[4] B. J. Coe and S. J. Glenwright, *Coord. Chem. Rev.*, 2000, **203**, 5–80.

[5] M. a. Halcrow, *Chem. Soc. Rev.*, 2013, **42**, 1784–1795.

[6] Z. M. Heiden and T. B. Rauchfuss, *J. Am. Chem. Soc.*, 2009, **131**, 3593–3600.

[7] G. Ramachandran, C. Ramakrishnan and V. Sasisekharan, *J. Mol. Biol.*, 1963, **7**, 95–99.

[8] D. Frishman and P. Argos, *Proteins: Struct., Funct., Bioinf.*, 1995, **23**, 566–579.

[9] A. D. Robertson and K. P. Murphy, *Chem. Rev.*, 1997, **97**, 1251–1268.

[10] H. A. Carlson and J. A. McCammon, *Mol. Pharmacol.*, 2000, **57**, 213–218.

[11] D. L. Mobley and K. a. Dill, *Structure*, 2009, **17**, 489–498.

[12] O. Keskin, A. Gursoy, B. Ma and R. Nussinov, *Chem. Rev.*, 2008, **108**, 1225–44.

[13] W. E. Stites, *Chem. Rev.*, 1997, **97**, 1233–1250.

Bibliography

[14] K. B. Mullis, *Nobel Lecture: The Polymerase Chain Reaction*, 1993.

[15] C. a. Hutchison, S. Phillips, M. H. Edgell, S. Gillam, P. Jahnke and M. Smith, *J. Biol. Chem.*, 1978, **253**, 6551–6560.

[16] M. Bucci, C. Goodman and T. L. Sheppard, *Nat. Chem. Biol.*, 2010, **6**, 847–854.

[17] J. Finkelstein, *Nature*, 2009, **460**, 813.

[18] C. J. Wilson, D. Apiyo and P. Wittung-Stafshede, *Q. Rev. Biophys.*, 2004, **37**, 285–314.

[19] C. Andreini, I. Bertini, G. Cavallaro, G. L. Holliday and J. M. Thornton, *J. Biol. Inorg. Chem*, 2008, **13**, 1205–1218.

[20] Y. Lu, N. Yeung, N. Sieracki and N. M. Marshall, *Nature*, 2009, **460**, 855–862.

[21] L. Rulísek and J. Vondrásek, *J. Inorg. Biochem.*, 1998, **71**, 115–127.

[22] K. Pisarewicz, D. Mora, F. C. Pflueger, G. B. Fields and F. Marí, *J. Am. Chem. Soc.*, 2005, **127**, 6207–6215.

[23] T. D. Stephens, C. J. Bunde and B. J. Fillmore, *Biochem. Pharmacol.*, 2000, **59**, 1489–1499.

[24] J. J. Berzelius, *Ann. Chim. Phys.*, 1836, **61**, 146.

[25] V. Smil, *Nature*, 1999, **400**, 415.

[26] M. R. Elsegood and D. A. Tocher, *Polyhedron*, 1995, **14**, 3147–3156.

[27] N. Uematsu, A. Fujii, S. Hashiguchi, T. Ikariya and R. Noyori, *J. Am. Chem. Soc.*, 1996, **118**, 4916–4917.

[28] R. Noyori and S. Hashiguchi, *Acc. Chem. Res.*, 1997, **30**, 97–102.

[29] R. Noyori, *Angew. Chem. Int. Ed.*, 2002, **41**, 2008.

[30] W. S. Knowles, *Angew. Chem. Int. Ed.*, 2002, **41**, 1998.

[31] Y. Chauvin, *Angew. Chem. Int. Ed.*, 2006, **45**, 3740–3747.

[32] R. R. Schrock, *Angew. Chem. Int. Ed.*, 2006, **45**, 3748–3759.

[33] R. H. Grubbs, *Angew. Chem. Int. Ed.*, 2006, **45**, 3760–3755.

[34] A. Suzuki, *Angew. Chem. Int. Ed.*, 2011, **50**, 6722–6737.

[35] E.-i. Negishi, *Angew. Chem. Int. Ed.*, 2011, **50**, 6738–6764.

[36] A. Gutteridge and J. M. Thornton, *Trends Biochem. Sci.*, 2005, **30**, 622–9.

[37] J. E. Debreczeni, A. N. Bullock, G. E. Atilla, D. S. Williams, H. Bregman, S. Knapp and E. Meggers, *Angew. Chem. Int. Ed.*, 2006, **45**, 1580–1585.

[38] G. E. Atilla-Gokcumen, N. Pagano, C. Streu, J. Maksimoska, P. Filippakopoulos, S. Knapp and E. Meggers, *ChemBioChem*, 2008, **9**, 2933–2936.

[39] F. W. Monnard, T. Heinisch, E. S. Nogueira, T. Schirmer and T. R. Ward, *Chem. Commun.*, 2011, **47**, 8238–8240.

[40] C. J. Reedy and B. R. Gibney, *Chem. Rev.*, 2004, **104**, 617–649.

[41] V. M. Krishnamurthy, G. K. Kaufman, A. R. Urbach, I. Gitlin, K. L. Gudiksen, D. B. Weibel and G. M. Whitesides, *Chem. Rev.*, 2008, **108**, 946–1051.

[42] K. Okrasa and R. J. Kazlauskas, *Chem. Eur. J.*, 2006, **12**, 1587–1596.

[43] J. Podtetenieff, A. Taglieber, E. Bill, E. J. Reijerse and M. T. Reetz, *Angew. Chem. Int. Ed.*, 2010, **49**, 5151–5155.

[44] H.-S. Park, S.-H. Nam, J. K. Lee, C. N. Yoon, B. Mannervik, S. J. Benkovic and H.-S. Kim, *Science*, 2006, **311**, 535–538.

[45] H. Robinson, M. C. Ang, Y. G. Gao, M. T. Hay, Y. Lu and A. H. Wang, *Biochemistry*, 1999, **38**, 5677–5683.

[46] S. V. Wegner, H. Boyaci, H. Chen, M. P. Jensen and C. He, *Angew. Chem. Int. Ed.*, 2009, **48**, 2339–2341.

[47] M. E. Wilson and G. M. Whitesides, *J. Am. Chem. Soc.*, 1978, **100**, 306–307.

[48] T. R. Ward, *Acc. Chem. Res.*, 2011, **44**, 47–57.

[49] P. J. Deuss, G. Popa, C. H. Botting, W. Laan and P. C. J. Kamer, *Angew. Chem. Int. Ed.*, 2010, **49**, 5315–5317.

[50] G. Roelfes and B. L. Feringa, *Angew. Chem. Int. Ed.*, 2005, **44**, 3230–3232.

[51] A. J. Boersma, B. L. Feringa and G. Roelfes, *Org. Lett.*, 2007, **9**, 3647–3650.

[52] B. V. Popp and Z. T. Ball, *J. Am. Chem. Soc.*, 2010, **132**, 6660–6662.

[53] R. P. Megens and G. Roelfes, *Chem. Eur. J.*, 2011, **17**, 8514–8523.

[54] H. Yamaguchi, T. Hirano, H. Kiminami, D. Taura and A. Harada, *Org. Biomol. Chem.*, 2006, **4**, 3571–3573.

[55] J. Collot, J. Gradinaru, N. Humbert, M. Skander, A. Zocchi and T. R. Ward, *J. Am. Chem. Soc.*, 2003, **125**, 9030–9031.

[56] M. Skander, N. Humbert, J. Collot, J. Gradinaru, G. Klein, A. Loosli, J. Sauser, A. Zocchi, F. Gilardoni and T. R. Ward, *J. Am. Chem. Soc.*, 2004, **126**, 14411–14418.

[57] C. Letondor, N. Humbert and T. R. Ward, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 4683–4687.

[58] M. Creus, A. Pordea, T. Rossel, A. Sardo, C. Letondor, A. Ivanova, I. Letrong, R. E. Stenkamp and T. R. Ward, *Angew. Chem. Int. Ed.*, 2008, **47**, 1400–1404.

[59] M. Dürrenberger, T. Heinisch, Y. M. Wilson, T. Rossel, E. Nogueira, L. Knörr, A. Mutschler, K. Kersten, M. J. Zimbron, J. Pierron, T. Schirmer and T. R. Ward, *Angew. Chem. Int. Ed.*, 2011, **50**, 3026–3029.

[60] A. Pordea, M. Creus, J. Panek, C. Duboc, D. Mathis, M. Novic and T. R. Ward, *J. Am. Chem. Soc.*, 2008, **130**, 8085–8088.

[61] J. Pierron, C. Malan, M. Creus, J. Gradinaru, I. Hafner, A. Ivanova, A. Sardo and T. R. Ward, *Angew. Chem. Int. Ed.*, 2008, **47**, 701–705.

[62] C. Lo, M. R. Ringenberg, D. Gnandt, Y. Wilson and T. R. Ward, *Chem. Commun.*, 2011, **47**, 12065–12067.

[63] V. Köhler, J. Mao, T. Heinisch, A. Pordea, A. Sardo, Y. M. Wilson, L. Knörr, M. Creus, J.-C. Prost, T. Schirmer and T. R. Ward, *Angew. Chem. Int. Ed.*, 2011, **50**, 10863–10866.

[64] T. K. Hyster, L. Knörr, T. R. Ward and T. Rovis, *Science*, 2012, **338**, 500–503.

[65] P.-O. Norrby and P. Brandt, *Coord. Chem. Rev.*, 2001, **212**, 79–109.

[66] A. Leach, *Molecular Modelling: Principles and Applications (2nd Edition)*, Prentice Hall, 2001, p. 784.

[67] P. Hohenberg and W. Kohn, *Phys. Rev.*, 1964, **136**, B864–B871.

[68] W. Kohn and L. J. Sham, *Phys. Rev.*, 1965, **140**, A1133–A1138.

[69] A. D. Becke, *J. Chem. Phys.*, 1993, **98**, 5648–5652.

[70] P. J. Stephens, F. J. Devlin, C. F. Chabalowski and M. J. Frisch, *J. Phys. Chem.*, 1994, **98**, 11623–11627.

[71] B. R. Brooks, R. E. Bruccoleri, D. J. Olafson, D. J. States, S. Swaminathan and M. Karplus, *J. Comput. Chem.*, 1983, **4**, 187–217.

[72] D. A. Case, T. A. Darden, T. E. I. Cheatham, C. L. Simmerling, J. Wang, R. E. Duke, R. Luo, R. C. Walker, W. Zhang, K. M. Merz, B. Roberts, S. Hayik, A. Roitberg, G. Seabra, J. Swails, A. W. Goetz, I. Kolossváry, K. F. Wong, F. Paesani, J. Vanicek, R. M. Wolf, J. Liu, X. Wu, S. R. Brozell, T. Steinbrecher, H. Gohlke, Q. Cai, X. Ye, M.-J. Hsieh, G. Cui, D. R. Roe, D. H. Mathews, M. G. Seetin, R. Salomon-Ferrer, C. Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko and P. A. Kollman, *Amber 12*.

[73] L. D. Schuler, X. Daura and W. F. van Gunsteren, *J. Comput. Chem.*, 2001, **22**, 1205–1218.

[74] W. L. Jorgensen, D. S. Maxwell and J. Tirado-Rives, *J. Am. Chem. Soc.*, 1996, **118**, 11225–11236.

[75] N. L. Allinger, R. A. Kok and M. R. Imam, *J. Comput. Chem.*, 1988, **9**, 591–595.

[76] S. J. Weiner, P. A. Kollman, D. T. Nguyen and D. A. Case, *J. Comput. Chem.*, 1986, **7**, 230–252.

[77] A. Vedani, *J. Comput. Chem.*, 1988, **9**, 269–280.

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

[79] A. D. Mackerell, M. Feig and C. L. Brooks, *J. Comput. Chem.*, 2004, **25**, 1400–1415.

[80] T. a. Halgren, *J. Am. Chem. Soc.*, 1992, **114**, 7827–7843.

[81] H. A. Lorentz, *Ann. Phys.*, 1881, **248**, 127–136.

[82] D. Berthelot, *C. R. Hebd. Acad. Sci.*, 1898, 1703–1855.

[83] N. Gresh, P. Claverie and A. Pullman, *Theor. Chim. Acta*, 1984, **66**, 1–20.

[84] N. Plattner and M. Meuwly, *Biophys. J.*, 2008, **94**, 2505–2515.

[85] N. Gresh, G. A. Cisneros, T. a. Darden and J.-P. Piquemal, *J. Chem. Theory Comput.*, 2007, **3**, 1960–1986.

[86] P. J. Steinbach and B. R. Brooks, *J. Comput. Chem.*, 1994, **15**, 667–683.

[87] P. P. Ewald, *Ann. Phys.*, 1921, **369**, 253–287.

[88] T. Darden, D. York and L. Pedersen, *J. Chem. Phys.*, 1993, **98**, 10089–10092.

[89] E. J. Corey and J. C. Bailar, *J. Am. Chem. Soc.*, 1959, **81**, 2620–2629.

[90] P. V. Bernhardt and P. Comba, *Inorg. Chem.*, 1992, **31**, 2638–2644.

[91] A. Ghosh, *Curr. Opin. Chem. Biol.*, 2003, **7**, 110–112.

[92] P. Comba, *Coord. Chem. Rev.*, 2003, **238-239**, 9–20.

[93] M. Zimmer, *Coord. Chem. Rev.*, 2009, **253**, 817–826.

[94] L. Hu and U. Ryde, *J. Chem. Theory Comput.*, 2011, **7**, 2452–2463.

[95] I. Tubert-Brohman, M. Schmid and M. Meuwly, *J. Chem. Theory Comput.*, 2009, **5**, 530–539.

[96] A. K. Rappe, C. J. Casewit, K. S. Colwell, W. A. Goddard and W. M. Skiff, *J. Am. Chem. Soc.*, 1992, **114**, 10024–10035.

[97] S. C. Hoops, K. W. Anderson and K. M. Merz, *J. Am. Chem. Soc.*, 1991, **113**, 8262–8270.

[98] D. V. Sakharov and C. Lim, *J. Am. Chem. Soc.*, 2005, **127**, 4921–4929.

[99] R. H. Stote and M. Karplus, *Proteins: Struct., Funct., Bioinf.*, 1995, **23**, 12–31.

[100] M. Schmid, E. S. Nogueira, F. W. Monnard, T. R. Ward and M. Meuwly, *Chem. Sci.*, 2012, **3**, 690.

Bibliography

[101] A. Vedani and D. W. Huhta, *J. Am. Chem. Soc.*, 1990, **112**, 4759–4767.

[102] G. Rossato, B. Ernst, M. Smiesko, M. Spreafico and A. Vedani, *ChemMedChem*, 2010, **5**, 2088–2101.

[103] V. J. Burton, R. J. Deeth, C. M. Kemp and P. J. Gilbert, *J. Am. Chem. Soc.*, 1995, **117**, 8407–8415.

[104] R. Deeth, *Coord. Chem. Rev.*, 2001, **212**, 11–34.

[105] R. J. Deeth, A. Anastasi, C. Diedrich and K. Randell, *Coord. Chem. Rev.*, 2009, **253**, 795–816.

[106] R. J. Deeth, N. Fey and B. Williams-Hubbard, *J. Comput. Chem.*, 2005, **26**, 123–130.

[107] C. E. Schäffer and C. K. Jø rgensen, *Mol. Phys.*, 1965, **9**, 401–412.

[108] R. J. Deeth, *Chem. Commun.*, 2006, 2551–2553.

[109] R. J. Deeth, *Inorg. Chem.*, 2007, **46**, 4492–4503.

[110] N. Gresh, *J. Comput. Chem.*, 1995, **16**, 856–882.

[111] N. Gresh, J.-P. Piquemal and M. Krauss, *J. Comput. Chem.*, 2005, **26**, 1113–1130.

[112] M. P. Hodges, A. J. Stone and S. S. Xantheas, *J. Phys. Chem. A*, 1997, **101**, 9163–9168.

[113] N. Gresh, S. A. Kafafi, J.-F. Truchon and D. R. Salahub, *J. Comput. Chem.*, 2004, **25**, 823–834.

[114] J. Antony, J.-P. Piquemal and N. Gresh, *J. Comput. Chem.*, 2005, **26**, 1131–1147.

[115] H. M. Senn and W. Thiel, *Angew. Chem. Int. Ed.*, 2009, **48**, 1198–1229.

164

[116] M. Svensson, S. Humbel, R. D. J. Froese, T. Matsubara, S. Sieber and K. Morokuma, *J. Phys. Chem.*, 1996, **100**, 19357–19363.

[117] M. Elstner, D. Porezag, G. Jungnickel, J. Elsner, M. Haugk, T. Frauenheim, S. Suhai and G. Seifert, *Phys. Rev. B*, 1998, **58**, 7260–7268.

[118] M. Elstner, T. Frauenheim, E. Kaxiras, G. Seifert and S. Suhai, *Phys. Status Solidi B*, 2000, **217**, 357–376.

[119] Q. Cui, M. Elstner, E. Kaxiras, T. Frauenheim and M. Karplus, *J. Phys. Chem. B*, 2001, **105**, 569–585.

[120] D. Hartsough and K. Merz, *J. Phys. Chem.*, 1995, **99**, 11266–11275.

[121] U. Ryde, *Curr. Opin. Chem. Biol.*, 2003, **7**, 136–142.

[122] G. N. Lewis, *J. Am. Chem. Soc.*, 1916, **38**, 762–785.

[123] J. N. Murrell, S. F. A. Kettle and J. M. Tedder, *The chemical bond*, Wiley, 1978.

[124] L. Pauling, *J. Am. Chem. Soc.*, 1931, **53**, 1367–1400.

[125] P. Atkins and L. Jones, *Chemical Principles: The Quest for Insight*, W. H. Freeman, 2001, p. 1024.

[126] D. Manthey, *Orbital Viewer*, 2004.

[127] L. Gagliardi and B. O. Roos, *Nature*, 2005, **433**, 848–851.

[128] D. M. Root, C. R. Landis and T. Cleveland, *J. Am. Chem. Soc.*, 1993, **115**, 4201–4209.

[129] T. Cleveland and C. R. Landis, *J. Am. Chem. Soc.*, 1996, **118**, 6020–6030.

[130] C. R. Landis, T. Cleveland and T. K. Firman, *J. Am. Chem. Soc.*, 1998, **120**, 2641–2649.

Bibliography

[131] T. K. Firman and C. R. Landis, *J. Am. Chem. Soc.*, 2001, **123**, 11728–11742.

[132] C. R. Landis, T. Cleveland and T. K. Firman, *J. Am. Chem. Soc.*, 1995, **117**, 1859–1860.

[133] L. Pauling, *Proc. Natl. Acad. Sci. U. S. A.*, 1976, **73**, 274–275.

[134] S. F. Sousa, P. A. Fernandes and M. J. a. Ramos, *Proteins: Struct., Funct., Bioinf.*, 2006, **65**, 15–26.

[135] S. Cooper, F. Khatib, A. Treuille, J. Barbero, J. Lee, M. Beenen, A. Leaver-Fay, D. Baker, Z. Popović and F. Players, *Nature*, 2010, **466**, 756–760.

[136] F. Khatib, F. DiMaio, S. Cooper, M. Kazmierczyk, M. Gilski, S. Krzywda, H. Zabranska, I. Pichova, J. Thompson, Z. Popović, M. Jaskolski and D. Baker, *Nat. Struct. Mol. Biol.*, 2011, **18**, 1175–1177.

[137] G. M. Morris, R. Huey, W. Lindstrom, M. F. Sanner, R. K. Belew, D. S. Goodsell and A. J. Olson, *J. Comput. Chem.*, 2009, **30**, 2785–2791.

[138] O. Trott and A. Olson, *J. Comput. Chem.*, 2010, **31**, 455–461.

[139] Schrödinger, *Glide*, 2012.

[140] G. Jones, P. Willett and R. C. Glen, *J. Mol. Biol.*, 1995, **245**, 43–53.

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

[142] B. Seebeck, I. Reulecke, A. Kämper and M. Rarey, *Proteins: Struct., Funct., Bioinf.*, 2008, **71**, 1237–1254.

[143] L. Verlet, *Phys. Rev.*, 1967, **159**, 98–103.

[144] J. Ryckaert, G. Ciccotti and H. Berendsen, *J. Comput. Phys.*, 1977, **23**, 327–341.

[145] D. Shaw, M. Deneroff and R. Dror, *Comm. ACM*, 2007, **51**, 91–97.

166

[146] B. Rogaski and J. B. Klauda, *J. Mol. Biol.*, 2012, **423**, 847–861.

[147] P. Ferrara, J. Apostolakis and A. Caflisch, *Proteins: Struct., Funct., Bioinf.*, 2002, **46**, 24–33.

[148] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J. Chem. Phys.*, 1983, **79**, 926–936.

[149] F. C. Bernstein, T. F. Koetzle, G. J. Williams, E. F. Meyer, M. D. Brice, J. R. Rodgers, O. Kennard, T. Shimanouchi and M. Tasumi, *Arch. Biochem. Biophys.*, 1978, **185**, 584–591.

[150] B. Amrein, M. Schmid, G. Collet, P. Cuniasse, F. Gilardoni, F. P. Seebeck and T. R. Ward, *Metallomics*, 2012, **4**, 379–388.

[151] C. Andreini, L. Banci, I. Bertini and A. Rosato, *J. Proteome Res.*, 2006, **5**, 196–201.

[152] C. Andreini, L. Banci, I. Bertini, S. Elmi and A. Rosato, *Proteins: Struct., Funct., Bioinf.*, 2007, **67**, 317–324.

[153] C. Andreini, L. Banci, I. Bertini and A. Rosato, *J. Proteome Res.*, 2008, **7**, 209–216.

[154] G. Debret, A. Martel and P. Cuniasse, *Nucleic Acids Res.*, 2009, **37**, W459–W464.

[155] R. Wang, Y. Lu, X. Fang and S. Wang, *J. Chem. Inf. Comput. Sci.*, 2004, **44**, 2114–2125.

[156] Y. Zhang and J. Skolnick, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 1029–1034.

[157] M. Saito, N. Takemura and T. Shirai, *J. Mol. Biol.*, 2012, 1–12.

[158] C. T. Supuran, A. Scozzafava and A. Casini, *Med. Res. Rev.*, 2003, **23**, 146–189.

[159] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kale and K. Schulten, *J. Comput. Chem.*, 2005, **26**, 1781–1802.

[160] J. M. Zimbron, T. Heinisch, M. Schmid, D. Hamels, E. S. Nogueira, T. Schirmer and T. R. Ward., *J. Am. Chem. Soc.*, 2013.

[161] F. H. Allen, *Acta Crystallogr B.*, 2002, **58**, 380–388.

[162] J. Huang, D. Häussinger, U. Gellrich, W. Seiche, B. Breit and M. Meuwly, *J. Phys. Chem. B*, 2012, **116**, 14406–14415.

[163] M. Schmid, T. R. Ward and M. Meuwly, *J. Chem. Theory Comput.*, 2013.

[164] M. Schmid, T. R. Ward and M. Meuwly, *in preparation*, 2013.

[165] K. J. Bowers, R. O. Dror and D. E. Shaw, *J. Comput. Phys.*, 2007, **221**, 303–329.

[166] B. Hess, C. Kutzner, D. van der Spoel and E. Lindahl, *J. Chem. Theory Comput.*, 2008, **4**, 435–447.

Curriculum Vitae

Personal Information

Maurus Schmid
Neuweilerstrasse 83
4054 Basel
E-mail: maurus.schmid@unibas.ch

Date of birth: 12.05.1983
Nationality: Swiss

Education

University of Basel, Basel, Switzerland

PhD Studies Chemistry   **September 2008 - February 2013**

- Dissertation Topic: "Computational Modelling of Metal-Mediated Protein-Ligand Interactions"
- Advisors: Thomas R. Ward, Markus Meuwly

MSc Chemistry   **March 2007 - June 2008**

- Master Thesis Topic: "Development and Testing of an Approach to Implement the *trans* Influence into a Force Field"
- Advisor: Markus Meuwly

BSc Chemistry   **October 2003 - March 2007**

Publications

A dual anchoring strategy for the localization and activation of artificial metalloenzymes based on the biotin-streptavidin technology. Zimbron, J. M.; Heinisch, T.; Schmid, M.; Hamels, D.; Schirmer, T.; Ward, T. R. *Journal of the American Chemical Society.* 2013.

Development of Broadly Applicable Force Field Parameters for d^6-piano stool complexes. Schmid, M.; Ward, T. R.; Meuwly, M. *Journal of Chemical Theory and Computation.* 2013.

Identification of two-histidines one-carboxylate binding motifs in proteins amenable to facial coordination to metals. Schmid, M.; Amrein, B.; Collet, G.; Cuniasse, P.; Gilardoni, F.; Seebeck, F. P.; Ward, T. R. *Metallomics: Integrated biometal science.* 2012, 4, 379-388.

Arylsulfonamides as inhibitors for carbonic anhydrase: prediction & validation.
Schmid, M.; Nogueira, E. S.; Monnard, F. W.; Ward, T. R.; Meuwly, M. *Chemical Science.* 2012, 3, 690-700.

Molecular Mechanics Force Field for Octahedral Organometallic Compounds with Inclusion of the Trans Influence. Tubert-Brohman, I.; Schmid, M.; Meuwly, M. *Journal of Chemical Theory and Computation.* 2009, 5, 530-539.

Publications in preparation

Simulation of proteins involving d^6-piano stool complexes. Schmid, M.; Ward, T. R.; Meuwly, M.