PŘÍRODOVĚDECKÁ FAKULTA UNIVERZITY KARLOVY V PRAZE
KATEDRA BIOCHEMIE

Modelling of the interaction of proteins and peptides with metal ions

Modelování interakce proteinů a peptidů s kovovými ionty

Diploma thesis

Bc. Ondrej Gutten
Supervisor: Doc. RNDr. Jan Konvalinka, CSc.
Tutor: RNDr. Lubomír Rulíšek, Ph.D.

Prague 2010

This diploma thesis has been elaborated by the joint efforts at the Institute of Organic Chemistry and Biochemistry under the supervision and tutorship of Doc. RNDr. Jan Konvalinka, CSc. and RNDr. Lubomír Rulíšek, Ph.D.

I declare I have worked independently and quoted all the sources of information.

Prague, September 1, 2010 .....

Acknowledgments

I would like to express my deep appreciation of all the time, patience, advice and endless effort my tutor, RNDr. Lubomír Rulíšek, Ph.D., has invested in me and the project we have undertaken. The same recognition goes to my supervisor, Doc. RNDr. Jan Konvalinka, CSc. for the trust he bestowed on me and for the opportunity to work and learn in a very stimulating environment.

I would like to thank all my colleagues for friendly and pleasant environment, especially to Martin Srnec for fruitful and inspiring discussions.

A special thanks goes to Marián Šuch for indirect, yet indispensable contribution to production of this work.

I have benefited greatly from the company and discussions with these and other people, whom I am unable to name all, but who deserve no lesser appreciation.

Contents

List of abbreviations

B3LYP	functional for DFT method
BSSE	Basis Set Superposition Error
CBS	Complete Basis Set
CCSD(T)	Coupled clusters including single and double excitations and triple excitations obtained with perturbation theory
COSMO	Conductor-like screening model
DFT	Density Functional Theory
HF	Hartree-Fock
ITC	Isothermal Titration Calorimetry
MALDI-TOF	Matrix-Assisted Laser Desorption/Ionization-Time Of Flight
MP2	Moller-Plosset second order perturbation method
MPPT	Moller-Plosset Perturbation Theory
PBE	functional for DFT method
PCM	Polarizable Continuum Model
PDB	Protein Data Bank
PES	Potential Energy Surface
RI	Resolution of Identity
SPE calculation	Single Point Energy calculation
TPSS	functional for DFT method

1 Introduction

1.1 Bioinorganic chemistry

Although most of metals occur in living organisms only in trace amounts, their presence is absolutely indispensable. Most of the twenty-five essential elements are metals[1]. They function in almost all biological processes - cellular respiration, signal transduction, enzymatic catalysis being the most prominent examples.[1]

Metals also play important role in medicine as imaging and therapeutic agents.[2] Examples include also non-essential elements, such as Li, Ba, Ga, In, Tc, Pt, Gd, Sm and others. The fact that mechanism of their action is often far from being understood, despite their wide-spread application, prevents from systematic improvement of these agents[2].

Utilization of transition metal properties, their maintenance and toxicity bestow ample attractive problems to be studied by inorganic chemistry.

1.1.1 Metalloproteins

Among various essential elements in biocatalysis, metalloproteins play a specific role by catalysing reactions that would not occur under physiological conditions. The presence of metal ions is thus crucial for the oxidation/reduction processes, electron transfer, spin-forbidden reactions and 'difficult reactions', such as N_2, O_2, C–H bond breaking.[3] These processes are intimately involved in the fundamental elements of life, e.g. respiration and photosynthesis. Furthermore, metal ions play essential role in promoting processes such as protein folding, electron transfer, energy transfer, intra- and intercellular signalling and many others. It is, therefore, not surprising that approximately one third of proteins are estimated to contain metal ions[4].

Enormous efforts, both experimental and theoretical, have been exerted to understand the metalloprotein structure and function, including the correlation between the two.[5,6,7] However, the experimental and theoretical methods exhibit certain limitations,

preventing any of them from providing a comprehensive and unambiguous picture of their catalytic action.

The ultimate goal of our ongoing efforts (in which the presented diploma thesis represents the first step) is the design of novel peptide sequences mimicking the active site of metalloproteins, which would be enabled by the unique set of computer programs developed recently in our laboratory (the illustration of using the program for the design of putative iron(III)/zinc(II) active site mimicking the function of purple acid phosphatase is depicted in Figure 1). We believe that the above-described efforts will result in the discovery of new concepts that govern the topology and folding of metalloproteins and their catalytic action.

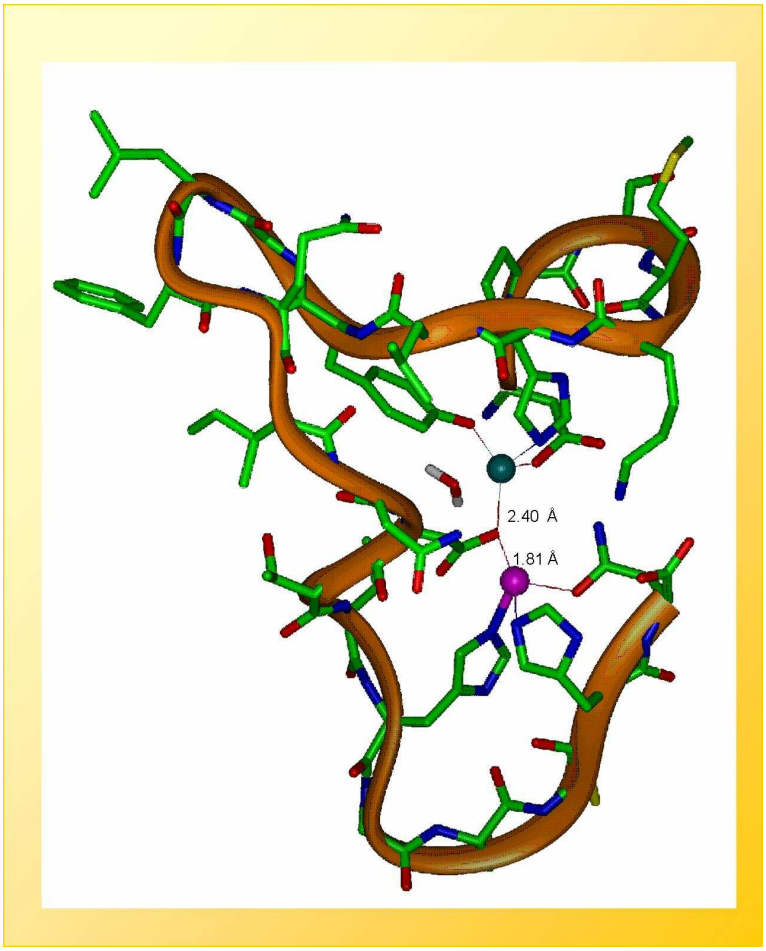


Figure 1: The putative structure of the *in silico* designed active site of purple acid phosphatase. Purple and dark green spheres represent Fe(III) and Zn(II). The most important bond distances are displayed. Backbone is in ribbon representation, while side-chains are displayed explicitly.[8]

The strategy leading to the above goals starts with studying the metal-ion selectivity by theoretical methods and design of the highly selective metal-binding sites, perhaps seeking to answer the intriguing question of why Nature selected various metal ions for performing specific functions can be better understood if we fully understand factors determining the selectivity of their binding in biomolecules.[9]

All the above issues represent central questions in the area of bioinorganic chemistry which is further discussed in the following section.

1.1.1.1 Function of metal ions in metalloproteins

Metal ions indulge in various roles, including catalysis, regulation and structure stabilization. Various metals possess unique properties that favour them in performing specific functions and this fact reflects in the measure of their exploitation by different organisms. For example, there is a significantly larger proportion of Fe-, Co-, Mn-binding proteins in archea and bacteria than in eukaryotes, while opposite is true for Zn ion[10]. The abundance of Zn domains[11], specifically zinc-finger motifs that are mostly associated with DNA binding and gene regulation, hints that this trend is connected to more complex organization of eukaryotic genome.

Regulatory metalloproteins can possess functional scaffolds that act as signal transducers, such as zinc-cysteine coordination environments, that influence other processes via controlled binding and release of zinc ion[12].

Metalloenzymes catalyse a wide variety of reactions, that could not otherwise occur under physiological conditions, such as reaction involving O_2, N_2, radical reactions, CH_4 formation, etc.[3]

A numerous group of metalloenzymes are oxidoreductases.[13] Metal ions in these enzymes are usually directly involved in the redox reaction, donating/accepting electrons from reactive species, but can also have additional function such as activating the substrate.[14] Metal ions in oxidoreductases are often bound by cofactors, which can fine-tune their redox potentials to fit various specific needs.[15]

In non-redox reactions, metal ion usually acts in electrostatic stabilization of intermediates or transition states or in destabilization of reactants, or helps with a proper substrate orientation. Activation of reactants involves withdrawal of electrons from ligands bound to the metal, inducing polarization of its bonds. The classical examples are polarization of P-O bond of ATP by Mg ions in kinases[16,17] or increasing the acidity of a substrate/cofactor and proton release.[18] Conversely, stabilization of intermediates and transition states exploits positive charge of metal ion to counterbalance local negative charge.[19] In fact, the two functions of stabilizing intermediates and activation of reactants are often coupled, thus increasing the metal's contribution to lowering the activation barrier.[15]

The example of the complex function that metalloproteins play in the nature is depicted in Figure 2. The enzyme that is depicted in the figure belongs to the class of multi-copper oxidases (MCOs). The MCOs are enzymes that couple the four-electron reduction of dioxygen in the trinuclear copper cluster with four one-electron oxidations of the substrates at the distant Cu-T1 site.[20]

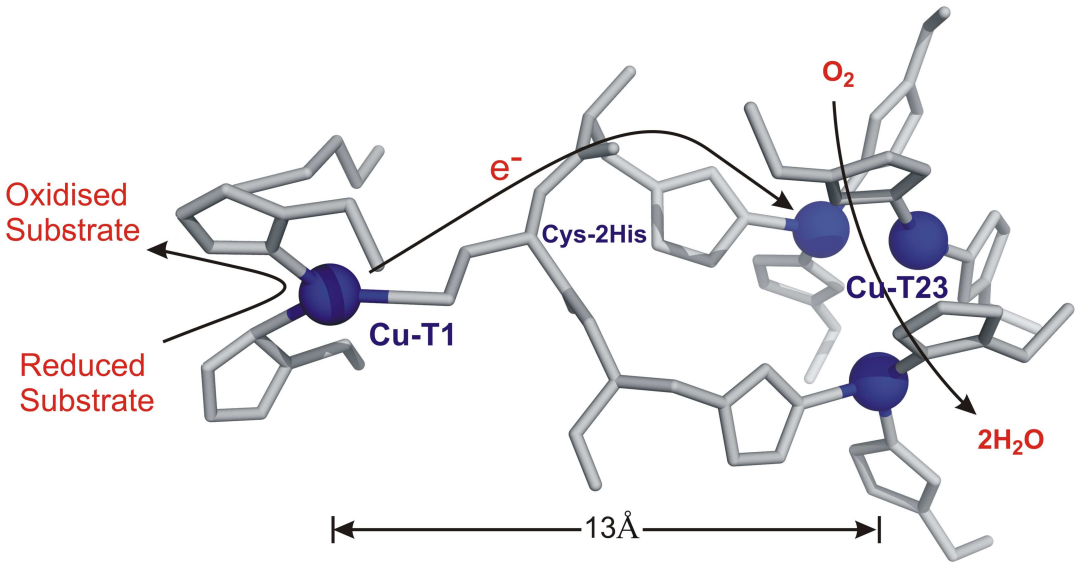


Figure 2: **Multi-copper oxidase active site.** Schematic drawing of the active site displays transport of electrons and catalytic reactions. Sidechains of aminoacidic residues that bind copper ions are displayed.

1.1.1.2 Metal uptake

The staggering fact that metalloproteins manage to pick their partner metal in a cell, although other metals are available, raises a question concerning the underlying mechanisms. The logical fact that metalloprotein needs to be stable with its metal partner does not imply its selection is purely on a thermodynamical basis. Even thermodynamically stable holoproteins are potential subjects for competition with metals that form less, yet comparably stable analogs. On the other hand, such proteins, or even proteins with lower affinities for target metals compared to other competing metal ions, could ensure stability by kinetically trapping the metal inside a protein. Nonetheless, these considerations suggest that the process of metalloprotein assembly is somehow mediated. Special group of proteins, called metallochaperones, are capable of binding a specific metal ion, transporting it to the target protein, recognizing it and finally passing it to the target protein.[21]

Rather few metallochaperons are known and it is thus assumed that most of the proteins acquire their metal partners from cellular pools and maintain them on thermodynamical basis. Only a few examples of mispopulation of metal ion have been reported,[22,23] although this could be due to difficulty in revealing such occasions.

Metal ions can act as toxic agents if present in inappropriate concentrations or oxidation states. Thus, organisms developed systems and mechanisms for sensing, transport and maintenance of homeostatic balance of metal ions. Response to metal concentration is mediated through metal sensors that can act as regulators of gene expression, both as activators or repressors.[24,25] Stronger interaction of sensors is found for metals farther up Irving–Williams series,[26] implying the lowest threshold value for metal ions such as Zn(II)[24] and Cu(II).[27] The selectivity is maintained through combination of factors, such as coordination geometry, coordination number and the nature of ligands or discriminating for specific charge of metal ion through number of negatively charged residues in the binding site.[27] The detection and uptake of specific metal ion is further governed by availability of different metal ions, implying that individual sensors can display variable sensitivity if observed, for example, in different organisms.[28,29] In other proteins, allosterical properties ensure that although various metal ions bind to the protein, but only the correct one triggers the conformational change that mediates the signal.[28,30,31] Sensors can induce expression of proteins that store, sequester or in other way regulate or

react to metal ion levels. One of the ways excess of heavy metals is dealt with is through synthesis of cysteine- or histidine-rich proteins that have strong metal-binding ability, such as metallothioneins[32] or Hpn proteins.[33,34]

Whether mere thermodynamical preference of one metal over others can be sufficient for design of a metal-selective protein is an interesting question and one of the aims of the work that will be succeeding this study.

1.1.2 Metal ion selectivity

Metal-ligand complexes have been extensively studied by inorganic chemistry. Nature of ligands, their geometry and orientation around the metal, solvent and electronic configurations all play major role in stability of the complex.[35] Following text focuses on ligands and their geometries, as these are the variables that will be varied throughout this work.

Geometry of the complex is closely related to a concept of coordination number, i.e. number of ligands bound to the metal. Coordination number is in turn determined mainly by the metal electronic configuration. Usually, the highest stability is achieved by completing its valence shell. The 18-electron rule helps to explain the stability of complexes with high covalency in the bonds, such as those containing for example carbonyls as ligands.[35] Its importance for bioinorganic chemistry where most of the bonds are ionic, with smaller degree of covalency, is somewhat limited. However, it still provides a qualitative explanation of the observation that different coordination numbers are preferred by different metals and their oxidation states.

Hard and soft acids and bases (HSAB) theory of Parr and Pearson[36] can provide a qualitative prediction and explanation of preference of transition metals for various ligands. Hard acids and bases are species with low polarizability and high charge-to-radius ratio (few representatives include Ti^{4+}, Co^{3+}, Cr^{3+} as Lewis acids and OH^{-}, F^{-}, Cl^{-}, NH_3, and CH3COO^{-} as Lewis bases). Conversely, soft acids and bases are easily polarizable, usually larger species with low charge. Common representatives are entire functional groups or atoms from higher periods of the periodic table (examples include Pt^{2+}, Pd^{2+},

Ag^{+}, Au^{+}, Hg^{2+}, Hg_2^{2+}, Cd^{2+} as soft Lewis acids and R_3P, SCN^{-}, I^{-} , and H_2S as soft Lewis bases).

The theory states that hard acids preferentially bind to hard bases and soft acids prefer soft bases. Hard acid – hard base interaction is mostly ionic. The combination of high charge and small size provides good conditions for strong Coulomb interaction. On the other hand, low polarizability presents poor condition for formation of covalent bonds. On the contrary, soft acid -soft base interaction has a large proportion of covalent bonding and dispersion interaction, due to good polarizability of both partners. Mixed soft-hard pairs cannot form any of these interactions as effectively and thus form comparatively less stable complexes.

1.1.3 Theoretical bioinorganic chemistry

1.1.3.1 Prediction of metal-ion selectivity

Computational modeling represents an indispensable tool in discovering fundamental physicochemical principles behind the chemical and biochemical processes.[37] One of the important biological phenomena is an uptake and binding of metal ions in biomolecules.[38] Since various metal ions play various roles in biological machinery, Nature fine-tuned the selectivity of various sites, such as metal-binding sites in metalloproteins for the specific ions.[39] Deciphering the mechanisms and factors behind the metal ion selectivity[40,41] is a highly desirable task which may ultimately lead to answering the question 'Why Nature selected specific metal ions for performing specific tasks?'[42]

Despite the accuracy of the computational methods for the description of the complex equilibria in biological fluids is far from the satisfactory,[43] it certainly complements experimental measurements in providing at least qualitative concepts concerning the details of the changes in the electronic and molecular structure upon metal complexation.[44,45] In most of the applications that aim at the accurate description of the biomolecular systems, the key feature is the ability of the method to describe both the intrinsic (gas-phase) energetics of the studied process as well as the process of solvation and desolvation of interacting species.[46,47]

Considering the fact that both solvation and desolvation of the charged species and

their *in vacuo* complexation energies are associated with large energetic changes (in the order of hundreds of kcal.mol^{-1} for divalent metal ions and negatively charged peptide species)[48] one may easily perceive that achieving the (bio)chemical accuracy of ~1-2 kcal.mol^{-1} (which translates into one unit in p*K* scale or one order of magnitude in the constants describing the equilibria in solution under normal pressure and temperature).[49]

There have been many studies addressing the problem of metal-ion selectivity from a computational and quantum chemical perspective.[50,51,52,53,54,55] These most often involved quantum chemical calculations of the small model of studied complexes, both in vacuo and polarized dielectric continuum in order to address the effects of the environment (e.g., solution, protein) and correlate the results with the phenomenological information obtained from the abundance of metal ions in the sites of metalloproteins. On the other hand, many attempts were made to address various equilibrium properties in the context of the full systems, which includes the calculations of reduction potentials[56] and protonation equilibria in proteins (e.g., pK_a values of protic functional groups).[57,58] However, to the best of our knowledge, there is not a computational method or protocol available that would enable us to reliably predict these properties with the accuracy challenging the experimental thermodynamic values acquired in condensed phase, i.e. within the accuracy of 1-2 kcal.mol^{-1} in free enthalpy (ΔG) value in solution.

1.1.3.2 Design of metalloproteins

Realization of the range of roles played by metal ions tempts to attempt designing a *de novo* metal-binding protein. Not only can it be a gateway to unlocking the unique chemical properties of metals for targeted exploitation in biological and/or technical utility, designing a peptide from scratch is an ultimate and thorough test of our knowledge of the field.

The simpler path of metalloprotein design tries to mimick native proteins. Most of the native proteins retain their stability even after undergoing multiple mutations and, hence, offer an excellent framework for studying its modification without raising concern about flawed stability of the new construct[59]. This is especially true for α-helical motifs, which are more independent compared to β-sheet structures, which were, not surprisingly, among the first successes of *de novo* design[60]. Analogue of O_2-binding haem proteins, that

retain affinity for O_2 similar to the natural globins, while possessing lower affinity for carbon monoxide, can serve as an exciting example.[61]

Further upside of mimicking a native protein when trying to design a new metalloprotein is that these scaffolds offer a versatile framework that is used by organisms themselves for a stunningly wide range of purposes. An example can be provided by β-barrel fold that is present in ca. 600 types of proteins with various functions.[62]

Numerous successful attempts have recently been made to alter the metal-binding properties of selected proteins with the aims of modifying protein function,[63,64,65,66,67,68] engineering novel metalloproteins or metal-binding sites in natural metalloproteins[69,70,71,72,73,74], or tuning metalloprotein or peptide specificity towards a particular metal ion[75,76,77,78,79].

On the other hand, design of completely new unnatural proteins opens a new world of possibilities, namely incorporation of unnatural amino acids. This idea is very appealing, leading to development of numerous techniques aiming at this goal.[80,81,82] Besides introducing new side-chains, it also offers an opportunity to modify protein backbone, for example by exchanging amidic peptide bond for an ester linkage.[83]

Modification of cofactors is also a hot topic, already with a number of successful applications. A trailblazing example is cast forward by Ru covalently attached to biotin, resulting in an artificial biocatalyst participating in reactions such as hydrogenation of alkenes, hydrogenation of ketones or carbon-carbon bond creation.[84,85,86,87,88]

Designing novel binding sites in peptides with high specificity for a particular metal ion is a highly attractive goal, namely for two reasons: it may provide a deeper understanding of the molecular basis for metal-ion specificity in protein and peptides on the one hand and binders for biomedical and technical applications on the other. These applications include the removal of metals from polluted environments, either by bacterial strains[89,90,91,92,93] or other biotechnological techniques;[94,95,96] the design of novel biosensors;[97,98,99] the redesign of proteins (providing new building blocks – maquettes);[100,101] or the design of new chelating compounds for medicinal chemistry.[39] Contribution to this field is one of the ambitious tasks for which the work presented in this thesis is but a first step.

1.2 Theoretical chemistry

1.2.1 Characterization and limitations

The goal of theoretical chemistry is not only to describe the underlying mechanisms behind the observed reality but also to predict the results of experiments. It means that mere existence of algorithms is not sufficient for a theory to be applicable, because existence of computational power able to solve problems in finite time is also required. This proved and still proves to be a significant problem in theoretical chemistry.

Analytical solutions of quantum mechanical problems are principally impossible even for the simplest chemical species. Introduction of numerical methods and simplified models transfers part of the problem into the aforementioned necessity for computational potency. These simplifications, even though far less accurate, cut down the demand on time drastically. However it remains the main limitation in the course of calculation.

Thus, selection of specific method is governed not only by its suitability for given problem but also by availability of computational power and time. This fact determines the nature of development of this scientific approach. *In silico* methods have also advanced significantly in the past few decades thanks to the vast progress in the field. Even though, contemporary theoretical chemistry may not appeal satisfactory, especially in the field of biochemistry where modelling of large and complex systems is required. However, the fact that the possibilities are to great extent delimited by hardware, which remains subject to unceasingly swift development, rather than by employed theory grants a great perspective to the field and many promises to the future.

1.2.2 Methods

Theoretical chemistry provides myriads of methods that differ not only in their approach, but more importantly in accuracy, computational cost and application. Choosing an appropriate method for dealing with a set-up problem is a paramount step.

Following paragraphs contain basic characterization and practical implications of methods relevant for this study. A more rigorous mathematical description and theoretical background can be found in the corresponding textbooks.[102,103,104]

1.2.2.1 Overview

We begin by introducing a term *ab initio* methods. *Ab initio* can be translated as 'from the beginning' (or 'from the first principles'), indicating that no empirical data are used. All of the properties (observables) of a given system are contained in its wave function, which is in general complex function that depends on spatial and spin coordinates of all elementary particles of the system. Non-relativistic wave function is obtained by solving Schrödinger equation:

$$i\hbar \frac{\partial}{\partial t} \Psi = \hat{H}\Psi \qquad (1)$$

The system is completely defined by its Hamiltonian ($\hat{H}$), which includes terms for electron-nucleus, nucleus-nucleus, electron-electron interaction, and external electric or magnetic fields interaction terms, if applicable. The solution of Schrödinger equation yields a set of eigenfunctions (Ψ_i) and a corresponding set of eigenvalues (E_i).

In most of chemical applications the time and spatial variables can be separated which leads to stationary Schrödinger equation which in the absence of external fields can be conveniently written as:

$$\hat{H}\Psi = (\hat{T}_n + \hat{T}_e + \hat{V}_{nn|} + \hat{V}_{ne|} + \hat{V}_{ee|})\Psi = E\Psi \qquad (2)$$

where $\hat{T}$ stands for kinetic energy, $\hat{V}$ for potential energy and subscripts *n* and *e* for nuclear and electronic parts, respectively.

Typically, kinetic motion of nucleus is significantly smaller than that of electrons, as nucleons (proton and neutron) are approximately 1800 times heavier than an electron, and can be separated from the electronic motion. This is called Born-Oppenheimer approximation[105]. The electronic part of Schrödinger equation then simplifies to

$$\hat{H}_e \Psi(R,r) = (\hat{T}_e + \hat{V}_{ne|} + \hat{V}_{ee|})\Psi(R,r) = E\Psi(R,r) \qquad (3)$$

which depends on positions of nuclei (*R*) only parametrically.

Born-Oppenheimer (or in general adiabatic) approximation is a plausible approximation in most of chemical application Its justifiability is ordained by the energy differences of individual electronic states of the system, but is usually considered as valid.[106,107]

1.2.2.2 Hartree-Fock[102]

Hartree-Fock method is central to quantum chemistry not only for being the simplest *ab initio* method for determining electronic structure, but also because it is the starting point for most of the more accurate (post Hartree-Fock, post-SCF) methods.

Principle

The HF method approximates the exact electron-electron interaction by an average potential field created by other electrons. Mathematically, this translates to expanding wavefunction into an antisymmetrized product of one-electron functions (atomic or molecular orbitals). This simplification allows for expression of the problem through a set of integro-differential equations that are each dependent on the coordinates of only one electron.

$$\hat{F}_i(\vec{1})\phi_i(\vec{1}) = \varepsilon_i\phi_i(\vec{1}) \qquad (4)$$

where ϕ_i is one one-electron function (orbital), $(\vec{1})$ marks dependence on three spatial and one spin coordinates of electron 1, ε_i is an energy of orbital ϕ_i and $\hat{F}_i$ is a Fock operator defined as

$$\hat{F}_i\phi_i = -\frac{1}{2}\nabla^2\phi_i - \sum_A \frac{Z_A}{r_{1A}}\phi_i + \sum_{j\neq i}\left[\int dx_2 \mid \phi_j(\vec{2}) \mid^2 r_{12}^{-1}\right]\phi_i - \sum_{j\neq i}\left[\int dx_2 \phi_j^*(\vec{2})\phi_i(\vec{2}) r_{12}^{-1}\right]\phi_j \qquad (5)$$

However, since the Fock operator depends on the orbitals on which it also operates, these problems need to be solved iteratively until the convergence in the energy and wave function is obtained. The final result is a set of orbitals (atomic or molecular) and their energies. Hartree-Fock ground-state wave function is constructed as Slater determinant consisting $N/2$ orbitals with lowest energy (where N is the number of electrons) in case of closed-shell electronic configuration.

As a consequence, Hartree-Fock is a *single-reference* wave function method. It is a variational method which implies that the energy obtained from the calculation is an upper bound to an exact energy. As the molecular orbitals are constructed from the atomic orbitals, or rather the so-called basis set functions, the calculated energy is dependent on the basis set used and should converge to the Hartree-Fock limit in case of hypothetically infinite basis set.

Limitations and scaling

Limitations of Hartree-Fock method are given by the simplifications applied – even with the (hypothetically) complete (infinite) basis set an exact energy cannot be reached. The difference between the exact non-relativistic energy within the Born-Oppenheimer approximation and the Hartree-Fock limit is called an *electron correlation energy*. This difference is present due to the simplification of exact two-electron interaction by an average potential that does not involve instantaneous electron-electron interactions (correlated movement of the many electrons), and due to a single-reference representation.

Hartree-Fock formally scales as $O(n^4)$ (n=number of basis functions) which rather stands at the lower end of the spectrum of *ab initio* methods.[104] With larger systems the actual scaling rather approaches $O(n^{2.7})$ with many efforts carried out to make it a linear scaling method.

Applications

Due to the inability to include electron correlation effects and to describe systems that cannot be described by a single-reference electronic configuration, Hartree-Fock wave

function (method) is nowadays used mostly as a starting point for more accurate methods (correlated wave function methods).

1.2.2.3 Perturbation theory

Principle[102]

As the name suggests, this method divides the system into zeroth order approximation and its (desirably small) perturbation. Through exact knowledge of the zeroth order approximation (i.e. Hamiltonian and its eigenfunctions), perturbation theory allows expression of all properties of the perturbed system as an infinite sum of contributions.

$$\begin{aligned}&\left(\hat{H}_0 + \lambda\hat{V}\right)\left(\left|\Psi_0\right\rangle + \lambda\left|\Psi_i^{(1)}\right\rangle + \lambda^2\left|\Psi_i^{(2)}\right\rangle + \ldots\right) = \\ &= \left(E_i^{(0)} + \lambda E_i^{(1)} + \lambda^2 E_i^{(2)} + \ldots\right)\left(\left|\Psi_0\right\rangle + \lambda\left|\Psi_i^{(1)}\right\rangle + \lambda^2\left|\Psi_i^{(2)}\right\rangle + \ldots\right)\end{aligned} \tag{6}$$

where $\hat{H}_0$ is the original Hamiltonian, $\hat{V}$ is a perturbation operator, λ is a dimensionless parameter set to unity, $E_i^{(n)}$ is n$^{\text{th}}$-order energy contribution, $\left|\Psi_0\right\rangle$ is an eigenstate of the original (non-perturbed) Hamiltonian and $\left|\Psi_i^{(n)}\right\rangle$ is n^{th}-order perturbation of this reference state that can be expressed in terms of eigenstates of unperturbed Hamiltonian.

The most common variant is Moller-Plesset perturbation theory (MPPT), which uses Hartree-Fock wave function as a zeroth order approximation. Zeroth and first order energy contributions sum up to the Hartree-Fock energy whereas higher order contributions constitute correlation energy.

Limitations and scaling

An expansion of MPPT into an infinite order of perturbation would yield exact ground state energy, but calculations of infinite number of expressions are technically impossible and the expansion needs to be terminated by including only the first *x* orders, *x* being typically 2-4. Choice of *x* is governed by the complexity of the studied system as the

computational cost quickly increases with the number of basis functions ($O(n^5)$ for $x = 2$ up to $O(n^9)$ for $x = 6$).[104] However, convergence of the MPx series is not always guaranteed which questions the accuracy of the truncated series. Nevertheless, comparison studies of MP methods with experiment or high level calculations show good agreement in most cases.[108],[109] By far, the most popular is the MP2 (second order Moller-Plesset) method since it represents a reasonable compromise between the accuracy and the computational cost.

Applications

Second order of Moller-Plesset perturbation theory (MP2) is the simplest correlation energy correction within the method and is used for medium sized systems. Higher orders contributions are usually used only for smaller systems.

1.2.2.4 Coupled clusters (CC) methods

Principle[104]

The CC wave function is expressed as a linear combination of Hartree-Fock ground-state wave-function and its various excitations using an exponential ansatz for the excitation operator. The CC method ensures size-consistency, i.e. linear scaling of energy with the number of electrons. The detailed description of CC expansion is beyond the scope of this diploma thesis.

Limitations and scaling

The main limitation of the CC theory is its high computational cost which depends on the level of excitations included in the CC ansatz. Thus, CCSD formally scales as $O(n^6)$, CCSD(T) scales as $O(n^7)$,[104] CCSDT as $O(n^8)$ and CCSDTQ as $O(n^{10})$.[104] Therefore, CCSD(T) methods are nowadays limited to the systems of 30-40 atoms, whereas CCSDT and CCSDTQ calculations are still considered as prohibitive for medium-sized molecules.

Applications

As mentioned above CCSD(T) is often considered as a reference method for systems with the single-reference ground electronic state as it mostly provides an excellent

agreement with experiment and can be used as predictive method in cases where experimental data are not available.[110,111]

1.2.2.5 Density functional theorem (DFT)

Principle [112]

All previously discussed methods were based on the wave functions (that uniquely determines properties of the system). Although many-electron wave function is intuitively constructed from one electron wave functions, it is a complex function of 4*N* spatial and spin variables (*N* is the number of electrons) which significantly complicates algebraic manipulation for larger systems.

An upside-down approach is presented in DFT methods that work with electron density rather than with coordinates of individual electrons. Electron density is function of only three spatial coordinates which simplifies comprehension as well as algebraic manipulation, while containing equivalent information. This equivalence is a consequence of the first Hohenberg-Kohn theorems state that the ground state properties (e.g., energy) of the system are uniquely determined by its electron density. The second Hohenberg-Kohn theorem is then an analogue of variational principle in wave function methods and it says that the energy obtained by inserting any trial density into the exact (unknown) energy functional is an upper bound to the exact energy.

If the exact form of the universal DFT functional is known it would provide us exact properties of the system. Unfortunately, this is not the case. In practice this problem has been solved by Kohn and Sham[113] who used a model of non-interacting electrons as the reference system. This leads to a simplification of the kinetic energy functional and a set of Kohn-Sham equations that formally resembles Hartree-Fock equations and can be solved iteratively. The price to be paid is that the terms describing electron-electron interactions (exchange and correlation effects) are lumped into the exchange-correlation functional whose form is also unknown. Nowadays, plenty of convenient DFT exchange-correlation functionals are available, derived quite often empirically (with certain boundary conditions to be satisfied) and calibrated against numerous thermodynamic, structural, and spectroscopic data.

DFT functionals

As mentioned above the exact form of universal DFT functional defining kinetic and correlation energy is unknown. Various forms based on approximations and/or parameters fitting have been elaborated, providing more or less satisfactory results. Strictly speaking, fitting of parameters to match experimental results casts use of such functionals into a domain of semi-empirical methods.

Three commonly used functionals, abbreviated as TPSS[114], B3LYP[115,116] and PBE[117], are used throughout this work.

Limitations and scaling

DFT scaling depends on the complexity of functional, but its scaling is in general similar to the scaling of the Hartree-Fock method. In practice, it is usually less than $O(n^3)$, especially for larger systems. DFT is therefore considered to be fairly cheap computational method.[118] However, there is an ongoing discussion concerning accuracy of DFT. Some studies show DFT is unable to correctly include long-distance dispersion interactions. This malady is often remedied by adding empirical dispersion term (DFT-D), which often improves the result significantly at no additional cost.[119]

Applications

Despite mentioned problems, DFT has been often shown to provide reasonable agreement with experiments and is nowadays considered as the best price/performance method in quantum chemistry.[120] It is often the only choice for quantum chemical study of large systems, including most of the biologically relevant systems. In the realm of metalloproteins, the use of DFT is almost exclusive and the accuracy obtained is usually satisfactory.[6]

1.2.3 Basis sets

Basis set is a set of tabulated one electron (spin-free) functions that is used to express wave function or electron density. Ideally, one would like to use the infinite (complete) basis set which is, however, not possible. For practical purposes, we need to

select a finite basis set that will provide a good approximation. In other words, we need to minimize the difference between an arbitrary function and its best approximation obtained from linear combination of functions from this basis set. However, this formulation of the task is not too helpful, since wave-function (or electron density) of the system is the unknown that we are trying to unfold.

Thus, development of a basis set is guided by chemical experience. The functions should have large enough values in the regions where the electron density is likely to be large and vice versa. However, this is not the only requirement. Computational cost for any method is dependent on the number of basis set functions. Choosing a large basis set has an upside of better description of the system, while smaller basis set lowers the cost of the calculation. Furthermore, it is imperative for these functions to have form that allows efficient calculations of integrals.[104]

Using finite basis sets poses another problem. As mentioned above, larger basis sets provide more accurate description of a system. Basis functions of any two atoms overlap and improve the description of the system even if these atoms do not interact, i.e. - in case of atom-centred basis functions more compact geometries are stabilized by the way the system is described. This stabilization, called basis set superposition error, is purely artificial and we seek to eliminate it. One of the ways to achieve this is by counterpoise correction method.[121] These undesired artificial contributions are obtained by SPE calculations of fragments of the given system which are, however, described by the basis functions of the entire system and these contributions are then subtracted from energy of the system.

1.2.4 Representation of water environment

Most of the biologically relevant molecules and systems are found in water environment. Therefore, representation of water environment is essential for accurate and relevant description of our systems. The straightforward approach of surrounding the system with a large number of water molecules, regularly applied in molecular dynamics simulations, is futile in quantum chemical calculations. Such an expansion of the system puts it beyond the reach of computability.

An alternative that will be used throughout the study is COSMO, conductor-like screening model[122]. The model treats solvent as a dielectric continuum that is present outside of the space occuppied by the studied system. This space is defined by Van der Waals radii of the atoms of the system. The model includes polarization of the continuum by solute. This model provides a reasonable approximation of water environment at incomparably lower cost.[123]

2 Aims of the Diploma Thesis

The primary aim of this work is to establish and, within limits of applied theory and available experimental data, justify strategy for *de novo* design of peptides or small proteins exhibiting high degree of metal ion selectivity.

Furthermore, we wish to provide refined input data based on this strategy that can be used for further development of novel peptides and as a core for database containing information about large number of metal-ligand complexes.

Further aim is to contribute to understanding the principles of the interactions and behaviour of selected transition metal ions with biologically relevant molecules and systems. This includes a careful benchmarking of the theoretical methods and computational strategies.

3 Methods and Systems

The scope of presented thesis can be divided into three areas:

- Justification of the selected computational method to be used for geometry optimization of metal-ligand complexes and calculation of their binding energies.
- Examination of the simplification of a system from metal-binding peptide in a water environment to a metal-ligand complex.
- Exploitation of this model in an ongoing effort to design peptides with desired selectivity properties.

The methodological issues pertinent to solving these problems are discussed in details below. This includes computational details and system setup.

3.1. Computational Details.

All quantum chemical calculations reported in this work were performed using TURBOMOLE 6.2 program.[124] The quantum chemical calculations were performed using the density functional theory (DFT) and correlated *ab initio* methods. Geometry optimizations were carried out either at the DFT level, employing Perdew-Burke-Ernzerhof (PBE) functional[117] or using the MP2 method. These DFT/PBE calculations were expedited by expanding the Coulomb integrals in an auxiliary basis set, the resolution-of-identity (RI-J) approximation.[125,126] For all geometry optimizations, the def-SVP basis set was employed on all atoms.[127,128]

The single-point DFT energies were calculated using the PBE,[117] B3LYP,[115,116] and TPSS[114] functionals. For these calculations the def2-TZVP basis set was employed on all atoms.[127] The *ab initio* reference energies were calculated using CCSD(T) method. In addition to the above basis sets, the correlation consistent aug-cc-pVDZ[129] and aug-cc-VTZ[130] basis sets were used. To allow for solvation effects, the conductor-like screening model (COSMO) method[131,132] was used with the dielectric constant corresponding to water ($\varepsilon_r = 80$)..

3.1 Towards Accurate Computational Modelling of Metal Ion Selectivity in Small Models of Protein Active Sites

In order to verify the reliability of DFT or MP2 method that were to be used for production calculations, five model complexes were built. The complexes are depicted in Figure 3.

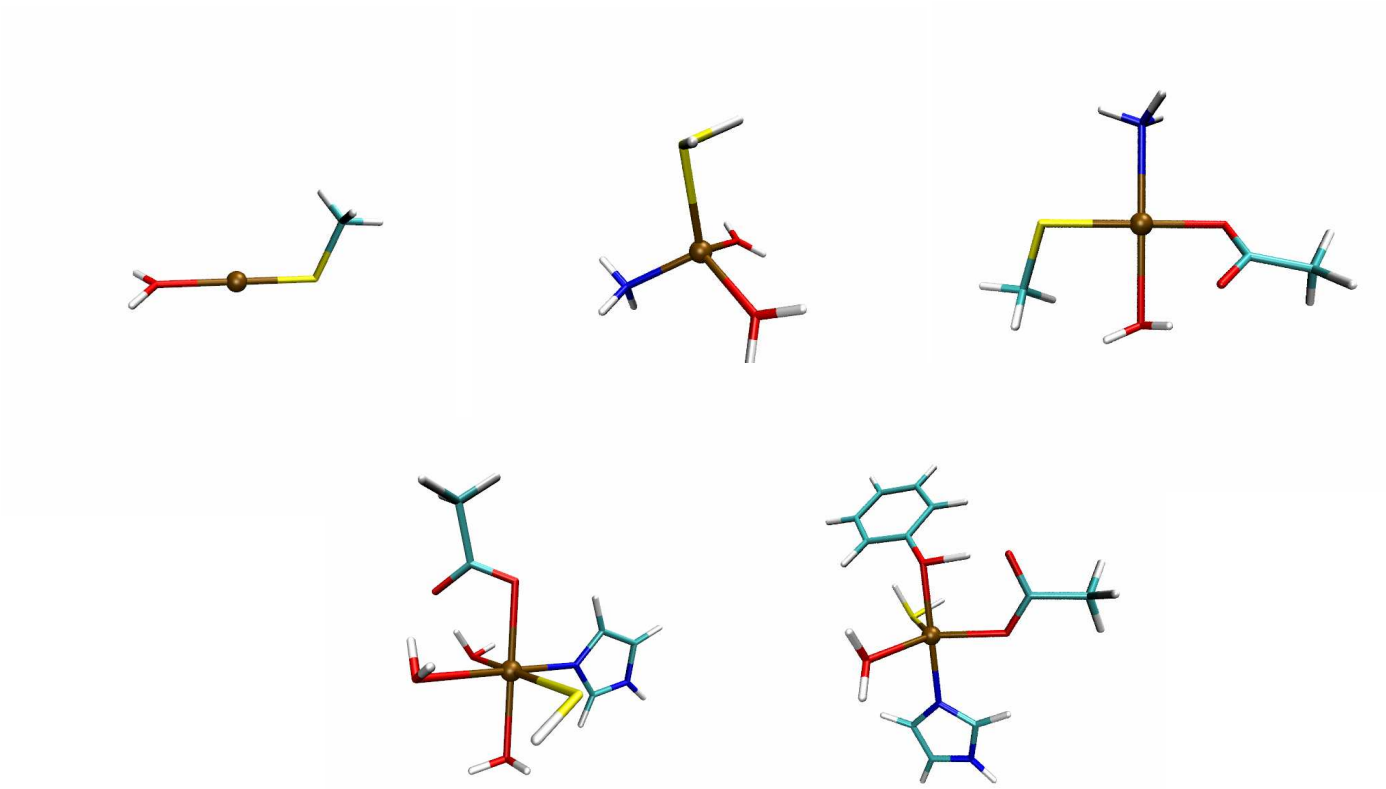


Figure 3: Model complexes. Upper left - $[M^{II}(CH_3S)(H_2O)]^{1+}$ in linear coordination geometry; Upper middle - $[M^{II}(H_2O)_2(H_2S)(NH_3)]^{2+}$ in tetrahedral coordination geometry; Upper right - $[M^{II}(CH_3S)(NH_3)(H_2O)(CH_3COO)]$ in square-planar coordination geometry; Lower left - $[M^{II}(H_2O)_3(SH)(CH_3COO)(Im)]$ in octahedral coordination geometry; Lower right - $[M^{II}(H_2S)(H_2O)(CH_3COO)(PhOH)(Im)]^{1+}$ in trigonal bipyramidal coordination geometry;

The choice of the combination of ligands is rather arbitrary but aims to include both charged and uncharged ligands and to range from small ligands (H_2O, H_2S) to the largest ones (PhOH).

Several sets of calculations were performed in an effort to compare differences in energy and geometries caused by use of different methods, basis sets, and use of COSMO model.

Each set consists of optimization part of the system and its reference and six single point energy (SPE) calculations:

-*system*:	ghost atoms: none	regular: all
-*system*:	ghost atoms: metal	regular: all but metal
-*system*:	ghost atoms: all but metal	regular: metal
-*reference system*:	ghost atoms: none	regular: all
-*reference system*:	ghost atoms: metal	regular: all but metal
-*reference system*:	ghost atoms: all but metal	regular: metal

where *system* stands for a complex in question and *reference system* stands for a complex with ligands substituted by water molecules. This scheme is set up to correct part of a basis superposition error.

The sets of calculations that were performed include combinations of different basis sets, geometry optimization methods, SPE calculation methods and environment (gas phase/COSMO model). Geometry optimizations were performed at DFT and MP2 level. Single point energy calculations were performed at DFT, MP2 and CCSD(T) level. Basis sets used for SPE calculations include def2-TZVP, aug-cc-pVTZ, aug-cc-PVDZ.

The purpose of the calculations was to investigate the influence of these factors on accuracy of predicted interaction energies, that we define as follows:

$$E_{\mathrm{i}}^{\mathrm{int}} = \left(E([\mathrm{ML_n}]) - E([\mathrm{G_M L_n}])\right) - \left(E([\mathrm{M(H_2O)_n}]) - E([\mathrm{G_M (H_2O)_n}])\right) \qquad (7)$$

where M denotes a metal ion, L denotes a ligand and G_x represents ghost atom with basis functions left from atom X. The overall charges of the complexes are not displayed in (7), since it varies (between 0 and 2+) in the model systems studied.

3.2 Predicting Stability Constants of Metal Ions in Peptidic Scaffolds from the First Principles

Before using the strategy for large-scale calculations, verification of the model is required. Since revelation of properties of metal selectivity is the major ambition this work pursuits, the goal of this verification is to calculate binding energies of selected sites with the selected metals and qualitatively reproduce these values in the simplified model.

3.2.1 Model peptides

Three small peptides (for which experimental values of binding energies are available as well) were used in this verification. Their initial geometries (see Figure 4 on the next page) were obtained from the program *build_peptide*.[133] It is important to mention that despite the studied peptides were synthesized and subject to experiments (MALDI-TOF, chelatometry, isothermal titration calorimetry, ITC), the structural information is missing. The efforts to crystallize the peptides or determined their structure by means of NMR were precluded by the fact that binding constants were probably too low for these experiments (~10^{-6}).

The peptides are labelled according to the amino acids expected to participate in the metal binding – CC(cysteine-cysteine), MM (methionine-methionine) and HHTC (histidine-histidine-threonine-cysteine)

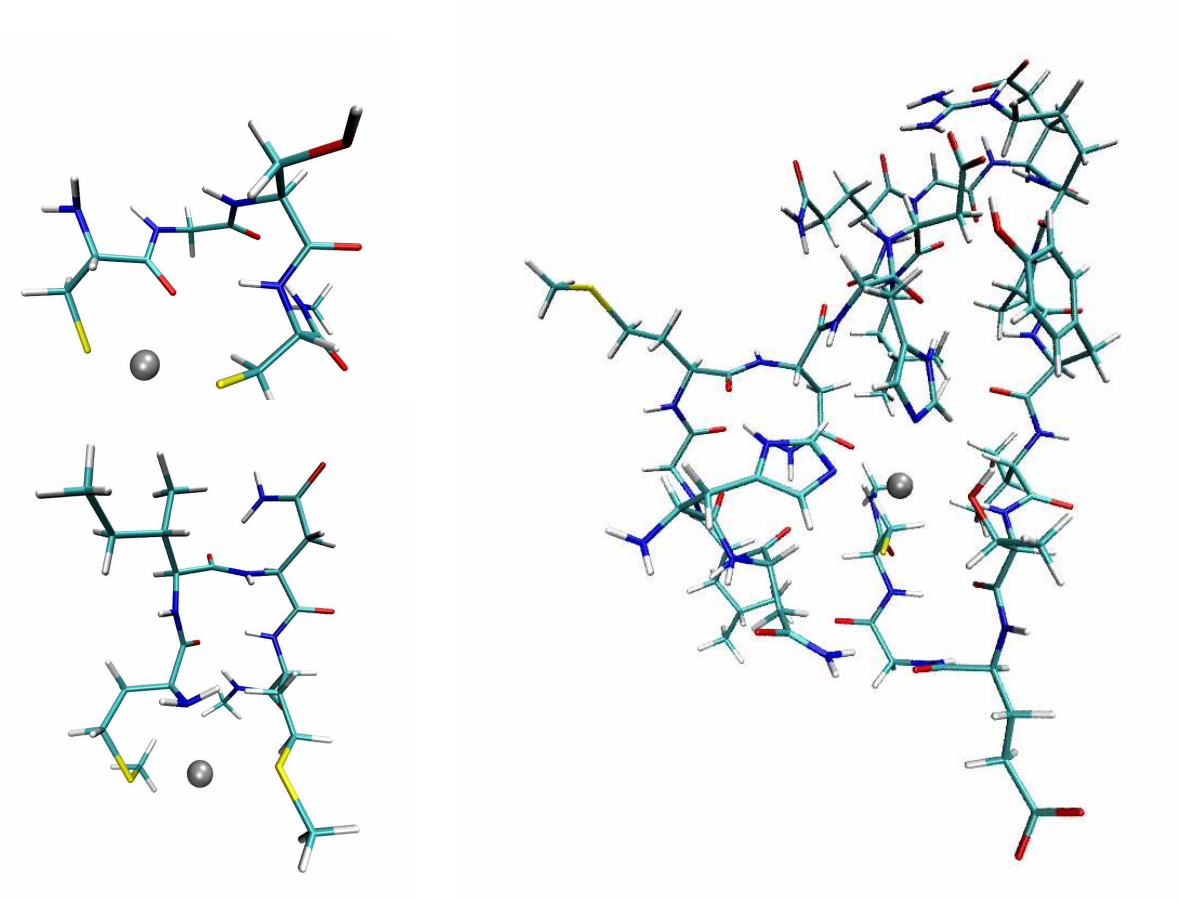


Figure 4.: Model peptides. Upper left - CGSC tetrapeptide (CC); Lower left - MINM tetrapeptide (MM); Right - HNLGMNHDLQGERPYVTEGC icosapeptide (HHTC)

Arbitrary metal ion is represented by a grey sphere.

3.2.2 Calculations of interaction energies

Model peptides

Set of calculations for computation of interaction energy consists of optimization part of the system and four single point energy (SPE) calculations:

-*system*:	ghost atoms: none	regular: all
-*system*:	ghost atoms: metal	regular: all but metal
-*reference system*:	ghost atoms: none	regular: all
-*reference system*:	ghost atoms: metal	regular: all but metal

where *system* stands for a complex in question and *reference system* stands for a complex of metal with six water molecules. This scheme is set up to correct part of a basis superposition error.

- All of the calculations of HHTC peptide were done in both forms of protonated (labelled HHTCp) and deprotonated (labelled HHTCd) threonine side chain.
- Structure optimizations of peptides were performed with and without constraints on geometry. In constrained optimizations individual ligand-metal-ligand angles were fixed.
- All of the calculations of model peptides were done with implicit solvent model provided by COSMO

Simplified first-shell model complexes

Simplified model complexed were constructed in the following manner:

1) Optimized geometry of the corresponding model peptide was considered

2) Coordinates of metal and binding aminoacidic sidechains were extracted and truncated in a manner described below. The missing bonding partner was substituted with a hydrogen atom.

3) In case of CC peptide the CO group of N-terminal cysteine binds to the metal and is included in the simplified model and is represented by a formamide $(NH_2)CHO$

In case of MM peptide the NH_2 group of N-terminal methionine binds to the metal and is included in the simplified model and is represented by an ammonia NH_3

The process is displayed in Figure 5 on the next page.

Set of calculations for computation of interaction energy consists of optimization part of the system and its reference and four single point energy (SPE) calculations:

-*system*:	ghost atoms: none	regular: all
-*system*:	ghost atoms: metal	regular: all but metal
-*reference system*:	ghost atoms: none	regular: all
-*reference system*:	ghost atoms: metal	regular: all but metal

where *system* stands for a complex in question and *reference system* stands for a complex of metal with six water molecules. This scheme is set up to correct part of a basis superposition error.

All of the calculations of simplified complexes were done both with implicit solvent model COSMO and in gas phase

Structure optimizations of simplified complexes were performed with and without constraints on geometry. In constrained optimizations individual ligand-metal-ligand angles were fixed.

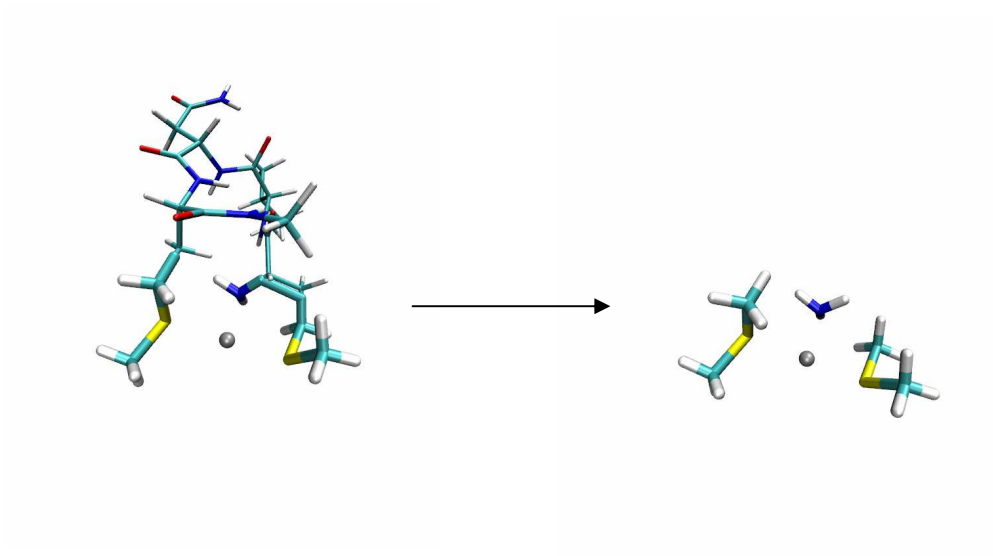


Figure 5: System simplification. Two methionine sidechains and terminal NH_2 are recognised as ligands and represented in the simplified model as dimethyl sulfides and ammonia, respectively. Position of atoms is not changed during the simplification but the resulting system is subjected to geometry optimization.

Reference systems:

Reference systems are complexes of corresponding metal with six water molecules preorganised in octahedral geometry and optimized with no constraints on geometry. Optimizations were performed both with implicit solvent model COSMO and in gas phase.

3.3 $[MX_n]$ complexes: introduction to combinatorial quantum chemistry

The final part comprises tremendous effort that has been undertaken to utilize the testing described in previous chapters. It is divided into three parts. The first part describes preparation of input files for all complexes that were to be investigated. Second part describes the calculations that have been performed with these complexes together with an outlook for the future. The final part expounds the process of construction of a peptide with desired metal coordination center.

3.3.1 Input file preparation

3.3.1.1 Selection of ligands

Out of 20 basic amino acids, 11 of them are capable of binding metal ions via their side chains – aspartic acid, glutamic acid, asparagine, glutamine, cysteine, lysine, histidine, methionine, serine, threonine, tyrosine. To consider all of the possible binding sites (with respect to metal-binding amino acids) we need to have a representative molecule for each one of them. However, since it is the side chain that should have decisive effect on binding of the metal the rest of the amino acid is excluded from the model. Instead, the side chain is terminated by a CH_3 group. Complete list of representatives is contained in Table 1 on the next page.

Table 1: List of ligands. *Lists ligands used to represent amino acidic side chains as metal-binders. The ligand is usually a truncated side chain, so that it possesses metal-binding atom and key atoms.*

Ligand	Represents	Ligand	Represents
	Glutamic acid	C_6H_5OH	
CH_3COO^-	Aspartic acid	$C_6H_5O^-$	Tyrosine
	Glutamine	CH_3SH	Cysteine
CH_3CONH_2	Asparagine	CH_3S^-	
$C_3H_4N_2$ (Imidazole)	Histidine	CH_3SCH_3	Methionine
CH_3OH	Serine	CH_3NH_2	Lysine
$CH_3CHOHCH_3$	Threonine		

Histidine can bind via a nitrogen atom of its imidazole ring. Both $N_{\delta 1}$ (N *pros*, abbreviated N_π, according to IUPAC nomenclature) and $N_{\varepsilon 2}$ (N *tele*, N_τ) binding modes are possible and should be included in the model. However, inclusion of each new representative into a set of possible ligands leads to a significant increase in the total number of complexes. Moreover, the difference between these two binding modes is likely to be unsubstantial or even imperceptible within the accuracy of the approach. Hence, inclusion of both modes would lead to a major increase in computational cost without actually increasing the information value contained in the results.

The very same argument applies to glutamic and aspartic acid being represented by a single ligand, as well as glutamine-asparagine couple.

On the other hand, some of the amino acids can acquire both charged and uncharged form. Since removal of hydrogen, otherwise bound to the ligating atom, has a major impact on electronic structure and overall behaviour of the ligand it is imperative to include both forms as two distinct ligands. This case applies for cysteine and tyrosine, which were also considered as cysteinate and tyrosinate.

Five geometries, eight metals and eleven ligands have been selected with an ambition to investigate all of the possible combinations of these variables. Even at first glance, it is obvious this will amount to a respectable quantity of complexes and even a mundane task of calculation set-up becomes a challenging quest.

Two steps need to be accomplished:

(i) acquiring a list of complexes to be studied;

(ii) obtaining initial guess of their geometries.

3.3.1.2 Constructing list of complexes

Constructing the list is not trivial. Due to symmetries of chosen geometries, simply writing out all possible values of variables (geometry, ligands, metals) would not suffice, as this list would contain redundant entries, which is undesirable. A following approach, that breaks the problem into several easily manageable sub-steps, has been used:

(1) A list of non-redundant templates is constructed. A template can be defined as a sequence of numbers, with each number defining a type of a ligand and the position of this number (in the sequence) defining ligand's position in the complex. If a template contains N different numbers, only numbers 1,2,...,*N* are used. We define *N* as cardinality of this template.

A new template is redundant if application of any element of symmetry (or their combination) yields a template that is identical to any of those already approved. A more illustrative clarification is presented in Figure 6 on the next page.

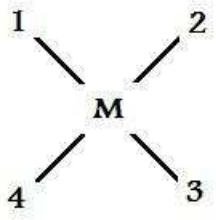
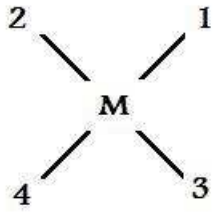
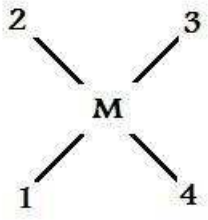
sketch of a complex			
template	1 2 3 4	2 1 3 4	2 3 4 1
reference number	1	2	3

Figure 6: Examination of template redundancy. Three sample templates and their schematic drawings are displayed. First number in the template defines upper left corner of the square, second number defines upper right corner, etc. in clock-wise direction.

Template 1 and 2 are non-redundant, since application of any combination of elements of symmetry to one template will not yield the other. On the other hand, templates 1 and 3 are redundant – template 3 is obtained from template 1 by a counter clock-wise 90° rotation. Therefore, only one of the two templates is selected.

(2) A single isomorphism is constructed between each *N*-combination from a set of ligands and set {1,2,...,*N*}. All non-redundant complexes are obtained by applying these isomorphisms to all non-redundant templates (naturally, only isomorphisms, whose domain's cardinality is equal to the cardinality of the template are applied).

(3) This procedure fails if some of the templates are omitted or the redundancy of the template is misjudged. To verify none of the templates are omitted, list of all possible (redundant) templates was prepared. Applying elements of symmetry to a given (possibly redundant) template yields an identical template to one and one only from the list of non-redundant templates, unless some have been omitted.

Applying elements of symmetry to individual (non-redundant) templates will not produce a different non-redundant template, unless its redundancy has been misjudged.

Usually, metal ions in proteins are not bound by more then 4 amino acid side chains.[9] To account for this observation, templates for trigonal bipyramidal and octahedral geometry consisted of four ligands plus one and two water molecules, respectively.

3.3.1.3 Obtaining an initial guess of the molecular geometry

First, the geometries of individual ligands were optimized. Second, the molecular geometry of octahedral complex $Zn[(H_2O)_5L]$ was optimized. The initial guess for the orientation of a ligand was guided by chemical intuition.

A program, called *rotate*, was written. It allows to construct specific complexes by using ligand-metal orientation obtained in the previous optimization. Thus, two out of three rotational degrees of freedom are defined. The remaining degree of freedom was determined by a rudimentary force field designed to prevent clash of ligands.

By this approach and applying the combinatorial algorithms described in the previous section, following numbers of metal complexes were constructed (using 11 ligands and H_2O):

66 for linear geometry
1001 for tetrahedral geometry
2211 for square planar geometry
4422 for octahedral geometry
7502 for trigonal bipyramidal geometry

Since only initial guess of geometry was available, the first obligatory step is geometry optimization. Optimizations were performed with constrained geometries by fixing donor-metal-donor angles. There are two reasons for choosing constrained optimization over fully relaxed (non-constrained) optimization. The first reason is that these complexes represent the metal-binding site of a protein, which is likely to stabilize the positions of individual side chains. Although the geometry can be distorted in a protein, and most likely it is, fixing the angles represents the stabilization of side chains better than no fixation at all. Second reason is that we are trying to investigate preference of studied metals for chosen geometries. Distortion, or even destruction, of predefined geometry would not provide information about the energetics we seek.

3.3.2 Quantum Mechanical Calculations of [MX_n] Complexes

Optimizations were performed in vacuum environment, i.e. *without* COSMO model. There are two reasons for choosing vacuum environment over COSMO model. One reason is technical. COSMO calculations are several times more consuming than gas phase calculation. Due to immense number of complexes that were to be investigated this saving is welcome. More importantly, in vacuum a complex prefers closely packed conformations, while in water environment more open structures are formed, as water can interact with ligands. The inside of a protein is often hydrophobic and tightly packed, so use of vacuum environment seems to be reasonable, at least for geometry optimization.

Optimized geometries of zinc complexes were used as initial guesses for optimization of the complexes of remaining metal ions. Again, there are two good reasons for this decision. Optimized geometry of an entire complex is a much better initial guess than mergence of optimized fragments, making its use economical as faster convergence can be expected. More importantly, the nearest local minimum is likely to be very similar to the starting point, which is a local minimum for complex with zinc. Thus, difference in energies will be mostly caused by exchange of metal, not by different conformations and interactions of ligands, thus yielding the requested information about metal ion selectivity.

Obtaining a convergence of molecular geometries for 8 x 15 202 = 121 616 is a highly untrivial task. Each one is a unique system with its own PES that may present unique challenges for achieving convergence. Although optimization algorithms are very efficient and successful in most circumstances, some cases require systematic intervention.

A script, called *feeder*, has been developed to undertake this challenge. It possesses two notable features. Firstly, it sets-up, submits, and regulates the calculations. Calculation set-up is dependent on the variables specific for the given calculation, such as overall charge of a complex, metal ion, geometry, etc. Regulation ensures effective exploitation of available computational power without overloading the computers or hard drives with data that cannot be processed in the imminent future, deletion of unnecessary data, and easy management.

Secondly, it checks the progress of submitted calculations. Successfully converged jobs are stripped of unnecessary data. In the opposite case, causes of failure are analysed

and, if within competence of the *feeder* script, an attempt to remove them is made based on the results of the analysis. If the script is unable to deal with the specific problem or the calculation has failed to converge despite multiple attempts at fine-tuning the parameters, all of the job's data are dumped for later analysis.

The script allows for a systematic progress in an ambitious quest with as little manual input as possible. Although not flawless, it enabled conveyance of a soaring amount of calculations that could hardly ever be managed without it.

3.3.3 Construction of new peptide

Another program, called *builder*, is still in development. Given the position of metal-binding amino acid side chains the program constructs a peptide chain with given amino acids in predefined positions.

The program works with a non-redundant PDB database of all known protein structures.

First it finds all peptide fragments that can connect two arbitrary ligands of a complex, i.e. it finds fragments that start with an amino acid identical to one ligand and ends with the other; few (usually three) more residues beyond these terminal ones are included (so called ’tails’) for reasons desribed in the next paragraph. Due to large number of such fragments only those satisfying the necessary conditions are saved. These conditions can be summed up into a following requirement: if the fragment is rotated correctly the terminal side chains have to be in desired positions, their orientation towards the position of the metal has to be correct and the peptide must not occupy space reserved for metal, other side chains or to interfere with the rest of the peptide.

In the next step, all combinations of these fragments are constructed. A scoring function has been devised to determine the quality of the new peptide. The scoring function takes into consideration deviations from prescribed side chain positions, angles and quality of tail overlaps. The idea behind the tail overlap is simple – in order to stabilize the position of ligating side chain it might be helpful to include amino acids from both directions of the peptide chain. Best combinations of fragments are merged and resulting peptides are printed out.

Although the idea of constructing a peptide by merging fragments is rather crude it has shown some success before[133]. The new construct can be subjected to geometry optimization in order to test its stability.

4 Results and Discussion

4.1 Towards Accurate Computational Modelling of Metal Ion Selectivity in Small Models of Protein Active Sites

4.1.1 The reference benchmark CCSD(T) calculations

For three smaller systems, $[M^{II}(CH_3S)(H_2O)]^{1+}$, $[M^{II}(H_2O)_2(H_2S)(NH_3)]^{2+}$, and $[M^{II}(CH_3S)(NH_3)(H_2O)(CH_3COO)]$ (Fig. 3a-c) and the corresponding perhydrated complexes: $[M^{II}(H_2O)_2]^{2+}$ and $[M^{II}(H_2O)_4]^{2+}$, it is possible to carry out the benchmark CCSD(T) calculations using a fairly large basis sets (up to aug-cc-pVTZ). For the current (benchmarking) purposes, the interaction energy is defined as the interaction of the bare metal ion with the 'pre-organized' binding site (obtained in the preceding molecular geometry optimization using the RI-PBE/def-SVP method). Three basis sets were used, viz def2-TZVP, aug-cc-pVDZ, and aug-cc-pVTZ. The results are summarized in Table 2 on the next page.

Several important observations can be made from the values presented in Table 2. Methodologically, the most important fact that can be noticed in Table 1, is a very good agreement between RI-MP2 values and the reference CCSD(T) values, in all three basis set studied. The difference in interaction energies computed in a given basis is usually 1-3 kcal.mol^{-1}. The notable exceptions are the linear complexes of Cu(II) and Fe(II). These will be investigated in more detail later. The small $\Delta E_{\text{MP2-CCSD(T)}}$ term also implies that one can attempt to obtain estimates of CCSD(T)/CBS values, as is routinely done in the calculations of interaction energies of weakly interacting systems. The MP2 (aDZ → aTZ → aQZ) calculations with the CCSD(T) (aDZ → aTZ) correction would serve to this purposes, though a fairly good estimate can be already provided by the estimate of MP2/CBS value. What is even more encouraging is that the errors are much lower once we adopt the definition of interaction energy as defined by (7).

Table 2: The interaction energies of studied metal ions with model binding sites, E_{int} = E(complex) – $E(M^{2+})$ – E(ligands) calculated using various *ab initio* and DFT methods. Equilibrium geometries were obtained using PBE/def-SVP method and constraining the systems to the given coordination geometries (i.e., fixing the L-M-L angles). All values are inkcal/mol.

coord. geom.	complex	MP2 TZV[a]	MP2 aDZ[b]	MP2 aTZ[c]	CCSD(T) TZV	CCSD(T) aDZ	CCSD(T) aTZ	PBE TZV	B3LYP TZV	TPSS TZV
LI[d]	$[ZnX_2]^{+}$[e]	-462.3	-468.4	-474.5	-461.4	-467.7	-474.3			
	$[ZnW_2]^{2+}$	-180.9	-185.7	-191.4	-179.1	-184.1	-190.4			
	$[CdX_2]^{+}$	-428.1	-429.1	-436.3	-425.4	-426.1	-432.9	-459.7	-450.3	-450.4
	$[CdW_2]^{2+}$	-145.4	-147.0	-152.4	-143.3	-144.9	-150.5	-163.2	-158.4	-158.2
	$[CuX_2]^{+}$	-506.8		-542.6	-490.1		-521.7	-566.2	-545.1	-553.9
	$[CuW_2]^{2+}$	-190.7		-199.9	-190.5		-200.8	-228.3	-209.5	-219.6
	$[FeX_2]^{+}$	-432.0		-426.1	-436.2		-432.7			
	$[FeW_2]^{2+}$	-163.3		-162.4	-163.0		-164.1			
TH	$[ZnX_4]^{2+}$[f]	-299.3	-305.3	-312.1	-296.7	-302.7	-310.5			
	$[ZnW_4]^{2+}$	-290.6	-291.9	-299.5	-288.7	-290.3	-299.0			
	$[CdX_4]^{2+}$	-249.6	-252.6	-259.6	-246.0	-248.8	-255.9	-272.6	-265.0	-264.8
	$[CdW_4]^{2+}$	-238.6	-237.8	-245.0	-235.5	-234.8	-242.3	-254.6	-250.9	-248.5
	$[CuX_4]^{2+}$	-299.8		-312.1	-301.8		-315.4	-353.0	-331.8	-343.6
	$[CuW_4]^{2+}$	-289.3		-297.9	-289.1		-299.7	-325.5	-311.6	-317.5
	$[FeX_4]^{2+}$	-271.2		-270.1	-271.0		-273.0			
	$[FeW_4]^{2+}$	-269.2		-266.1	-268.9		-268.9			
SQ	$[ZnY_4]$[g]	-732.8	-730.7	-738.1	-732.2	-730.0	-738.6			
	$[ZnW_4]^{2+}$	-281.8	-282.8	-289.8	-280.0	-281.5	-289.5			
	$[CdY_4]$	-669.2	-665.4	-673.1	-666.9	-662.8	-670.7	-695.1	-685.2	-688.1
	$[CdW_4]^{2+}$	-234.3	-233.0	-240.0	-231.3	-230.2	-237.5	-250.5	-246.5	-245.1
	$[CuY_4]$	-744.0			-751.6			-803.8	-781.7	-795.7
	$[CuW_4]^{2+}$	-301.0		-310.2	-301.5		-312.6	-337.9	-324.9	-332.0
	$[FeY_4]$									
	$[FeW_4]^{2+}$	-262.1			-262.0					

[a] def2-TZVP basis set

[b] aug-cc-pVDZ basis set

[c] aug-cc-pVTZ basis set

[d] LI… linear, TH… tetrahedral, SQ…square planar coordination geometry

[e] $[MX_2]$ stands for $[M^{II}(CH_3S)(H_2O)]^{1+}$ complex, $[MW_2]$… $[M^{II}(H_2O)_2]^{2+}$

[f] $[MX_4]$ stands for $[M^{II}(H_2O)_2(H_2S)(NH_3)]^{2+}$ complex, $[MW_4]$… $[M^{II}(H_2O)_4]^{2+}$

[g] $[MY_4]$ stands for $[M^{II}(CH_3S)(NH_3)(H_2O)(CH_3COO)]$ complex, $[MW_4]$… $[M^{II}(H_2O)_4]^{2+}$

In this equation the perhydrated complex in the given coordination geometry is used as the reference state. If we use the interaction energies as the central quantity that determines the metal ion selectivity, we can notice that the agreement between MP2 and CCSD(T) values drops down below 1 kcal.mol^{-1}, in most cases.

The calculations also nicely reproduce well known trends in the binding of studied metal ions (much higher binding energies for copper and mercury, the former is in agreement with Irving-Williams series of stability constants), preference of Cu(II) for square planar geometry (in comparison with the tetrahedral geometry).

At first glance, DFT values seem to be significantly off the values calculated by CCSD(T) or MP2. More careful examination, however, reveals that the shift is systematic, to a large extent. This can be seen if we apply equation (7) for calculation of interaction energy. Originally, differences in energy of complexes, i.e. values listed in Table £1, are in the range of 15-40 kcal.mol^{-1}. Upon applying (7), i.e. $E([MX_n]) - E([MW_n])$., the differences between DFT and CCSD(T) predicted values drop to 0-25 kcal.mol^{-1}. Significant differences in performance of individual functionals are encountered as well. Largest differences are found for the PBE functional. The TPSS functional performs much better and is only slightly surpassed by B3LYP, for which differences of 0-15 kcal.mol^{-1} are found.

These values are still quite large. Although it is tempting to use DFT for calculation of interaction energies with DFT, it has to be remembered that the 15 kcal.mol^{-1} value is an accuracy limit of the presented approach for the DFT method. Although the shift can be systematic and does not necessarily interfere with the goal of predicting metal ion selectivity, it disqualifies DFT from precise prediction of interaction energies.

The situation is similar for the more complicated octahedral and trigonal bipyramidal model complexes (Table 3on the next page). Agreement between MP2 and DFT improves after application of equation (7) but remains non-negligible and unsatisfactory. Although E_{int} predicted by MP2/aTZ and DFT/B3LYP for Cd-OH complex is identical, it is more of a solitary exception than a general rule.

These values further confirm that DFT is not appropriate for predicting highly accurate interaction energies, at least in the approach adopted in this study.

Table 3: Energies of octahedral and trigonal bipyramidal complexes. The interaction energies of studied metal ions with model binding sites, E_{int} = E(complex) – $E(M^{2+})$ – E(ligands) computed by using various *ab initio* and DFT methods. Equilibrium geometries were obtained using PBE/def-SVP method and constraining the systems to the given coordination geometries (i.e., fixing the L-M-L angles). The BSSE was accounted for using the counterpoise method of Boys and Bernardi. All values are in kcal.mol^{-1}.

coord. geom.	complex	MP2 aDZ[a]	MP2 aTZ[b]	PBE TZV[c]	B3LYP TZV[c]	TPSS TZV[c]
OH[d]	$[ZnX_6]^{+}$[e]	-736.7	-745.2	-755.0	-747.0	-751.8
	$[ZnW_6]^{2+}$			-374.9	-373.3	-371.6
	$[CdX_6]^{+}$	-669.6	-678.0	-693.8	-685.1	-688.3
	$[CdW_6]^{2+}$	-300.6	-308.5	-317.7	-315.6	-312.2
	$[CuX_6]^{+}$	-737.9	-745.7	-794.40	-772.7	-788.5
	$[CuW_6]^{2+}$	-359.9	-368.5	-365.1	-384.2	-390.2
TP	$[ZnX_5]^{+}$[e]	-538.0	-546.5			
	$[ZnW_5]^{2+}$			-343.2	-340.9	-339.0
	$[CdX_5]^{+}$	-481.5	-489.1	-495.4	-489.7	-490.2
	$[CdW_5]^{2+}$	-270.6	-278.1	-287.5	-284.6	-281.9
	$[CuX_5]^{+}$	-533.2		-601.4	-580.6	-595.4
	$[CuW_5]^{2+}$	-330.8	-339	-367.4	-354.3	-361.2

[a] def2-TZVP basis set
[b] aug-cc-pVDZ basis set
[c] aug-cc-pVTZ basis set
[d] OH… octahedral, TP… trigonal bipyramidal,
[e] $[MX_6]$ stands for complex $[M^{II}(H_2O)_3(SH)(CH_3COO)(Im)]$, $[MW_2]$… $[M^{II}(H_2O)_6]^{2+}$
[f] $[MX_5]$ stands for complex $[M^{II}(H_2S)(H_2O)(CH_3COO)(PhOH)(Im)]^{1}$, $[MW_4]$… $[M^{II}(H_2O)_5]^{2+}$

4.1.2 Comparison of geometry optimization at different levels of theory

As can be inferred from Tables 2 and 3, the interaction energies obtained using MP2 and DFT methods differ with the former ones being closer to the reference CCSD(T) calculations. The question naturally arises as to how accurate are equilibrium geometries obtained using DFT methods in comparison with the equilibrium MP2 geometries. For large-scale calculations, or the calculations of larger systems, the DFT/RI-PBE method is at least one order of magnitude faster than RI-MP2 method. Therefore, if it can be shown that the equilibrium geometries obtained using RI-PBE method are sufficiently accurate, the efficient and reasonably economic protocols such as the MP2/TZV//RI-PBE/def-SVP can be safely for sufficiently accurate modelling of the metal ion affinities in biomolecular sites.

To answer the question three different sets of geometries were studied.

For the first set an initial guess ("ig") was optimized with MP2 method (ig/MP2). For the second set an initial guess, identical with the previous one, was optimized with DFT-RI-PBE/def-SVP method (ig/DFT). For the third set DFT-optimized structures were optimized with MP2 method (DFT opt/MP2).

SPE calculations were undertaken for all these systems at MP2 level with aug-cc-pVTZ/aug-cc-pVTZ-PP basis. E_{int}, as defined previously in (7), are presented in the Table 4 below together with relative RMSD values and energy differences.

Table 4: the comparison of various geometries optimized with DFT and MP2 methods. *RMSD values between corresponding complexes of interest ML_n (*upper line*) and reference MW_n systems (lower line - italics) are presented.*

metal	coord.	E_{int}[f] starting geom. / optimization method			ig / MP2 vs ig / DFT		ig / DFT vs *DFT opt. / MP2*		ig / MP2 vs DFT opt. / MP2	
ion	**geom.**	***ig / MP2***	***ig / DFT***	**DFT opt. / MP2**	**ΔE_{int}**	**RMSD**	**ΔE_{int}**	**RMSD**	**ΔE_{int}**	**RMSD**
Cd	**LI**[a]					0.042		0.048		0.007
		-285.1	-283.8	-285.1	-1.3	*0.026*	1.3	*0.023*	0.0	*0.003*
	TH[b]					0.057		0.039		0.030
		-14.8	-14.6	-14.8	-0.2	*0.039*	0.3	*0.029*	0.0	*0.026*
	SQ[c]					0.150		0.155		0.043
		-441.7	-433.1	-442.0	-8.6	*0.015*	9.0	*0.015*	0.3	*0.005*
	OH[d]					0.126		0.122		0.011
		-378.0	-369.5	-378.1	-8.4	*0.009*	8.6	*0.015*	0.1	*0.007*
	TP[e]					4.066		0.266		4.078
		-224.0	-211.2	-214.8	-12.9	*0.020*	3.6	*0.015*	-9.6	*0.016*
Cu	**LI**[a]					0.035		0.035		0.017
		-341.5	-342.6	-341.6	1.1	*0.469*	-1.0	*0.005*	0.1	*0.003*
	TH[b]					0.996		0.359		0.845
		-14.8	-14.1	-14.5	-0.7	*0.122*	0.4	*0.121*	-0.4	*0.008*
	SQ[c]					2.110		0.921		2.353
		-451.8	-439.4	-452.5	-12.4	*0.132*	13.0	*0.133*	0.7	*0.005*
	OH[d]					0.312		0.317		0.026
		-385.1	-377.3	-385.9	-7.9	*0.149*	8.6	*0.151*	0.8	*0.014*

[a] LI stands for $[M^{II}(CH_3S)(H_2O)]^{1+}$ complex,
[b] TH stands for $[M^{II}(H_2O)_2(H_2S)(NH_3)]^{2+}$ complex,
[c] SQ stands for $[M^{II}(CH_3S)(NH_3)(H_2O)(CH_3COO)]$ complex,
[d] OH stands for $[M^{II}(H_2O)_3(SH)(CH_3COO)(Im)]$ complex,
[e] TP stands for $[M^{II}(H_2S)(H_2O)(CH_3COO)(PhOH)(Im)]^{1+}$ complex,
[f] Defined in (7)

The LI systems are very simple, containing only a metal ion and two ligands, in this case water and methanethiolate, offering little opportunity for overestimation or underestimation of different interactions. Thus, even very different optimization methods are unlikely to produce significantly distinct results. This expectation is confirmed by the findings. Energy differences of 1.3 and 1.1 kcal.mol^{-1} for Cd and Cu, respectively, are

inconsequential. Geometries are almost indistinguishable, as can be seen from very low values of RMSD. It is noteworthy, however, that upon return to MP2-optimization in DFTopt/MP2 set both energy differences and RMSD values (between LI systems of this set and of ig/MP2 set) drop almost to zero.

The situation is similar in TH systems. The system is slightly more complicated, containing metal ion and four ligands, which, in this case, are very simple uncharged ligands (H_2O, NH_3, H_2S). For Cd ion, just as in previous case, energy differences and RMSD values are diminutive and smallest for systems that are both optimized by MP2 method. However, situation is little more interesting with Cu ion. Although energy differences remain insignificant, the RMSD values are quite large. A more careful examination of the systems shows that the main differences come from different orientations of ligands – even though, the energies remain almost unchanged. This fact suggests that these orientations can vary without perceivable impact on overall interaction energy. This premonition is further encouraged by the observation that even after returning to MP2 for optimization, the calculated RMSD value (between ig/MP2 and DFTopt/MP2) remains quite large.

SQ systems are even more peculiar. For Cd ion the interaction energy of DFT optimized structure is significantly (ca. 9 kcal/mol) higher (less negative) than for both MP2 optimized geometries. The surprising fact is that RMSD values between these systems is rather small. Although it is quite possible that DFT optimized geometry is not a local minimum on MP2 energy evaluated PES, the presented differences are beyond expectation and we have no rigorous explanation for this insinuated discrepancy.

For Cu ion, both the energy differences and the RMSD values are quite large for comparison of DFT optimized and MP2 optimized systems. This seems to be an illustrative example of different energy evaluation by different methods and a resulting shift in geometry. A closer look reveals differences, mainly in orientations of larger ligands. Comparison of MP2 optimized geometries is noteworthy as well. The energy is almost identical, but the RMSD value is towering. This curiosity is caused by rotation of single ligand by almost 180°. Thus, the complexes are almost mutual mirror images, with very similar energetics. The alignment procedure, however, uses least-square method and thus aligns the structures, in this case, very inefficiently. Should we use a different alignment method, we could overlay the structures so that displacement of individual atoms would be close to zero – except for the atoms of rotated ligand.

Situation in OH systems is similar to the previous one. For Cd ion, RMSD values are quite low, but the energy differences between DFT and MP2 optimized methods are significant (ca. 8.5 kcal/mol). A return to MP2 optimization in DFTopt/MP2 Cd-OH system causes the differences to vanish. The situation is less disturbing for Cu ion, since the RMSD values are larger and the energy differences are more justifiable.

For TP systems, even MP2 optimized systems show significant differences both in energy and geometries, a typical example of different local minima found by starting from different initial geometries. Despite the large difference of ig/MP2 and ig/DFT, good agreement between ig/DFT and DFTopt/MP2 shows that DFT produced reasonable geometry that is close to MP2-evaluated PES minimum.

To conclude, we may say that in most cases DFT optimization provided geometries are similar to their MP2-optimized counterparts both in structure and energy. In a non-negligible number of cases, however, the differences are apparent. In some cases, energy differences are relatively large despite rather faint differences in geometry. Thus, if aforementioned protocol MP2/TZV//RI-PBE/def-SVP is to be used, error of up to 15 $kcal.mol^{-1}$ can be expected. The error is generally larger for more complex systems.

MP2-optimization with starting geometry taken from DFT-optimized systems yields results very similar to MP2-optimized systems that start with a mere initial guess of a geometry. A DFT-preoptimization with a subsequent MP2-optimization might be a bearable compromise between accuracy and computational cost.

4.1.3 Solvation effects (PCM/COSMO calculations).

Gas phase interaction energies are, in most cases, in the order of hundreds of $kcal.mol^{-1}$. This fact is not surprising, as reference system is composed of metal ion and water molecules, while studied complexes contain charged residues. Strong Coulomb interaction between this negatively charged fragments and positively charged metal ion is the reason for such large interaction energies. This applies for all (studied) metal ions, although the actual strength of the interaction varies slightly. On the other hand, TH systems, that contain no charged ligands have interaction energies of 10-20 $kcal.mol^{-1}$.

In water environment, implicitly represented in our calculations by polarizable continuum of COSMO model, this interaction is significantly weakened. Highly polar water molecules can provide stabilizing interaction for both negatively and positively

charged species. Thus, part of the charge cloud is “expended” on interaction with water. As a result, the ligand, even though still carrying formal charge, offers significantly weaker interaction and the overall interaction energy is much lower. This trend can be inferred from Table 5.

Table 5: Comparison of interaction energies calculated in gas phase and COSMO. SPE values of DFT-optimized structures at both DFT and MP2 level. Data for Cu-TP is not available.

metal ion	coord. geom.	charged ligands	method	Gas phase	COSMO
Cd	**LI**[a]	**1**	DFT	-291.9	-60.5
			MP2	-285.1	-55.3
	TH[b]	**0**	DFT	-14.1	-14.5
			MP2	-14.8	-13.9
	SQ[c]	**2**	DFT	-438.7	-85.6
			MP2	-441.7	-84.4
	OH[d]	**2**	DFT	-369.5	-48.8
			MP2	-378.0	-53.4
	TP[e]	**1**	DFT	-205.1	-35.8
			MP2	-224.0	-41.6
Cu	**LI**[a]	**1**	DFT	-335.5	-79.6
			MP2	-341.5	-84.4
	TH[b]	**0**	DFT	-20.2	-19.5
			MP2	-14.9	-14.2
	SQ[c]	**2**	DFT	-456.9	-85.7
			MP2	-451.8	-81.3
	OH[d]	**2**	DFT	-388.9	-63.0
			MP2	-385.2	-49.5
	TP[e]	**1**	DFT	-226.7	-41.5
			MP2		

[a] LI stands for $[M^{II}(CH_3S)(H_2O)]^{1+}$ complex,
[b] TH stands for $[M^{II}(H_2O)_2(H_2S)(NH_3)]^{2+}$ complex,
[c] SQ stands for $[M^{II}(CH_3S)(NH_3)(H_2O)(CH_3COO)]$ complex,
[d] OH stands for $[M^{II}(H_2O)_3(SH)(CH_3COO)(Im)]$
[e] TP stands for $[M^{II}(H_2S)(H_2O)(CH_3COO)(PhOH)(Im)]^{1+}$
[f] Defined in (7)

Upon introducing water environment, the interaction energy is lowered by a different amount, depending on the precise electronic structure of a given system. However, it can be claimed that the larger the number of charged ligands the larger the change of interaction energy, as larger amount of Coulomb interaction is lost due to interaction with surrounding water. This claim is consistent with the presented data.

The values are virtually unchanged for TH systems, which contain no charged ligands. This would tempt to conclude that the change in amount of Coulomb interaction is the only significant change that has impact on the interaction energy of a given system. This idea, however, has to be dismissed, as other ligands do interact with their environment as well through dipole-dipole, dipole-induced dipole or dispersion interactions. These interactions cause geometries optimized in vacuum environment to be more compact than those in water environment, as these, although weak, interactions are usually favourable and no better partners are present. However, when the system is surrounded with water molecules, these can compete for interaction with ligands. Large dipole moment of water can often provide stronger interaction than most of the ligands. Additionally, more open geometries offer more space for these interactions. Hence, it can be expected that introducing water environment can have a major impact on orientations of ligands and its overall geometry. The reason for utter indifference of interaction energies in TH systems to the environment might lie in the simple structure of its ligands (two H_2O molecules, H_2S and NH_3) that offers little opportunity to a more complex interaction, such as effect of more or less compact geometries that has been outlined above. Moreover, the interaction energy of this complex seems to have little dependence on the orientation of these ligands, as has been discussed in section 3.2

The interaction energies are much lower in the condensed phase and are thus closer to experimental values that are in the range of 0-20 kcal/mol. However, these values are still as high as 90 kcal/mol. It seems that COSMO model fails to fully supplement water environment and, thus, underestimates the stabilization provided by water molecules. A straightforward solution of surrounding the complex with another layer of water molecules could remedy the situation but, in return, introduces further problems. Specifically, introducing large number of water-water interactions impacts energy of the complex and eclipses the original issue. Since different geometries are studied it is impossible to ensure that the orientations of water molecules, and hence overall water-water interaction, are unchanged. A more thorough approach of sampling the phase space is well beyond the scope of this study.

4.2 Predicting Stability Constants of Metal Ions in Peptidic Scaffolds from the First Principles

4.2.1 Approximating the Complex Peptide Binding Site by Its First-Sphere Representation

We investigate the simplification of a metal-binding protein to a complex containing only the first-sphere ligands – amino acid side chains ligating the metal ion (and possibly backbone NH_2 or CO functional groups), and its effect on interaction energies (define in (7)). We do this by comparing the systems with identical environment (i.e. those optimized and calculated with COSMO model, as protein systems were not calculated in gas phase) and restrictions on geometry optimization (either no constraints or fixed donor-metal-donor angles). Calculated values of interaction energies for these systems are summarized in Tables 6 and 7.

Table 6: The interaction energies of whole peptides calculated with COSMO model. All values are in kcal.mol^{-1}.

Geom. constraints[a]	system[b]	MN	FE	CO	NI	CU	ZN	CD	HG	$E^{int}_{max} - E^{int}_{min}$
free	CC	8.6	3.4	1.1	4.4	-32.2	-8.0	-29.1	-80.5	89.1
	MM	87.5	88.1	83.1	91.3	51.3	74.6	52.5	11.8	79.5
	HHTCp	-27.6		-39.0	-32.4		-43.5	-51.1	-84.2	56.6
	HHTCd	-65.3	-75.1	-83.4	-71.7	-102.3	-80.1	-83.9	-112.0	46.7
fixed	CC	11.8	7.7	5.4	8.9	-25.7	-1.4	-28.1	-80.7	92.5
	MM	87.8	89.2	87.6	89.6	77.4	78.4	51.7	10.8	78.7
	HHTCp	-15.1				-60.6	-26.7	-42.2	-84.0	68.8
	HHTCd	-58.3				-108.3	-71.1	-74.4	-105.5	50.0

[a] … free – no constraints on geometry, fixed – fixed ligand-metal-ligand angles
[b] … model peptides described in chapter 3.2.1

Table 7: The interaction energies of simplified systems derived from whole peptides calculated with COSMO model. All values are in kcal.mol^{-1}.

Geom. constraints[a]	system[b]	MN	FE	CO	NI	CU	ZN	CD	HG	$E^{int}_{max} - E^{int}_{min}$
free	CC	1.0	-5.5	-10.6	-4.9	-45.9	-14.7	-34.9	-87.3	88.3
	MM	69.7	67.2	63.2	68.3	24.8	53.6	31.7	-7.7	77.4
	HHTCp	-16.3	-24.6	-28.8	-21.6	-48.8	-33.3	-40.0	-65.4	49.0
	HHTCd	-64.0	-75.1	-80.4	-76.3	-100.4	-79.4	-81.0	-116.1	52.1
fixed	CC	4.7	-0.6	-6.1	-2.7	-33.3	-12.7	-34.6	-83.7	88.4
	MM	72.8	69.9	69.3	72.3	33.6	64.2	36.8	-3.7	76.5

a … free – no constraints on geometry, fixed – fixed ligand-metal-ligand angles
b … simplified systems derived as described in chapter 3.2.2

The absolute values of complexation energies for original and simplified systems are different. This can be expected as the whole peptide provides more interaction partners than simplified system derived from it. Moreover, slightly different orientations of ligands are present. However, the range of these differences is quite narrow. For more illustrative display we define simplification energies as:

$$E^{sim}_{i} = E^{int}_{1st,i} - E^{int}_{orig,i} \qquad (8)$$

where $E^{int}_{orig,i}$ is a interaction energy metal ion in the full peptide and $E^{int}_{1st,i}$ is interaction energy of simplified system derived from its original as described in Computational Details section.

For constrained systems, where only the effect of different number of interaction partners takes place, simplification energies range from 7.1 to 14.3 for HHTCp, 6.9 to 11.4 for HHTCd, -11.5 to -3.0 for CC, and -19.4 to -14.1 for MM with a single exception of -43.8 for Cu; all values are in kcal.mol^{-1} (the reason for this large deviation is discussed in the next paragraph). This implies that outer layer does influence interaction energy but with limited regard to metal ion (4.6 to 8.6 kcal.mol^{-1} differences). For unconstrained systems,

where both effects are present, these values are slightly different (10.2 to 18.8 for HHTCp, -4.6 to 3.1 for HHTCd, -13.6 to -5.8 for CC and -26.4 to -17.8 for MM; in kcal.mol^{-1}) but the dependence on the metal ion remains similar (7.7 to 8.7 kcal.mol^{-1} differences). These numbers would suggest that differences in complexation energies of less than ~10 kcal.mol^{-1} are not reliably reproduced within this simplification. However, complexation energies span in the range of 40-90 kcal.mol^{-1} and therefore, some information concerning the potential selectivity is retained in the simplification.

Large value of simplification energy for Cu in constrained MM system is present due to non-consistent simplification of a protein to a first-layer complex. Simplified MM systems consist of two dimethyl sulfides (representing methionine side chain) and ammonia (representing terminal NH_2 functional group). The problem occurs with representation of this terminal NH_2 group. In all systems, except for the one containing Cu, the NH_2 group is bound to a metal. This favourable interaction is absent in the whole protein MM-Cu system. However, in the simplified representative of MM-Cu system this group is bound in the same way as in all other systems and represents a favourable interaction. Thus, the simplification energy for this system is ca. 25 kcal.mol^{-1} more negative than for other systems in this series. If this group is removed completely in the simplified system, we obtain E^{int}_{Cu} = 86.4 kcal/mol and E^{sim}_{Cu} = +8.9 kcal.mol^{-1} which is different from other values as well. Clearly, the correct description lies somewhere between the two values. Since this is a unique exception it is excluded from further discussion.

Despite the pessimistic outlook drawn by the estimated accuracy of this simplification the actual ordering of metal ion affinities in the studied systems for various metal ions is surprisingly well preserved (Figure 7 on the next page).

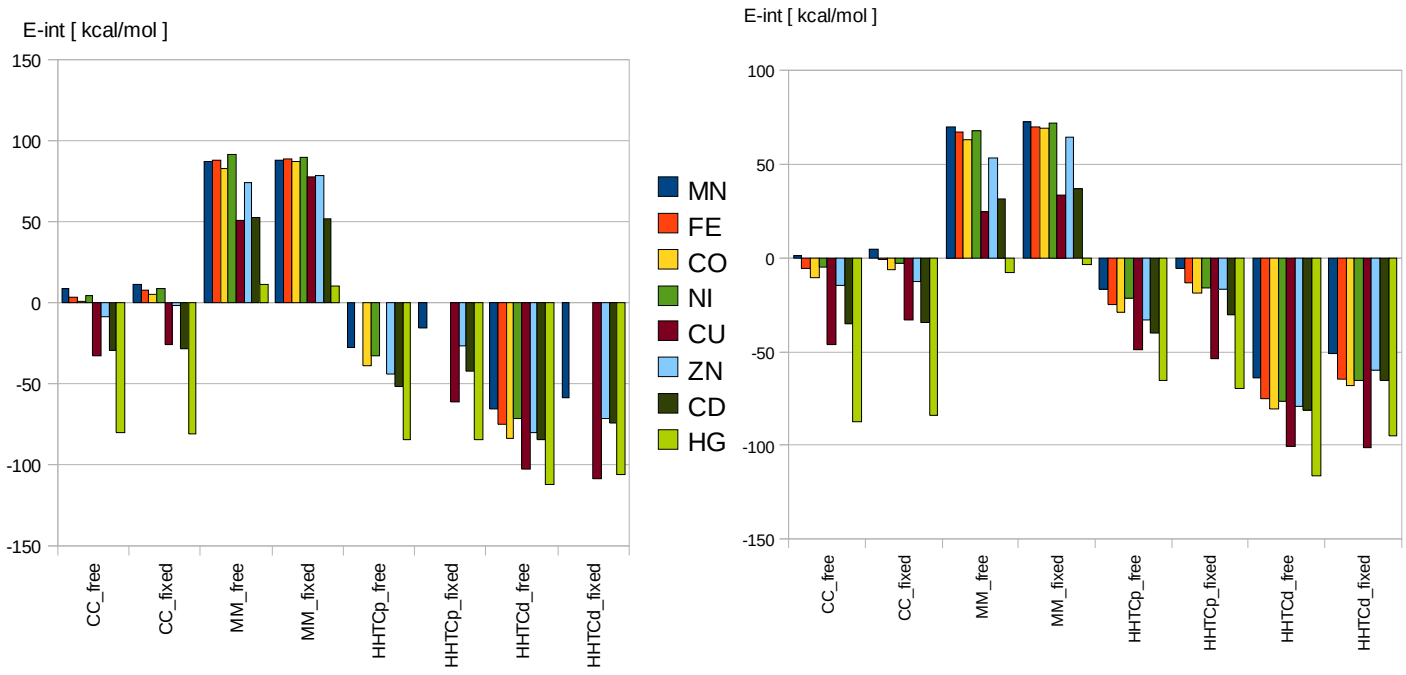


Figure 7. Relative affinities of systems for various metal ions. Interaction energies of eight divalent metal ions with studied systems are displayed.

Constrained systems

For CC system, relative order of affinities is preserved in the simplification. The single exception of Ni and Fe, where the order is swapped, shows difference in complexation energies of ~2 kcal.mol^{-1}.

For MM system, affinities for Mn, Fe, Co and Ni all lie within range of 2 kcal.mol^{-1}. Their relative order is different in the simplified systems but all these values lie in 3.6 kcal.mol^{-1} wide range. The overall order of affinities of Zn, Cd, Hg and the remaining group is preserved in the simplification. Cu is excluded from the discussion due to reasons addressed earlier.

In HHTCp, the order is preserved for all metals, where values are available (Hg, Cd, Zn, Cu, Mn). In HHTCd, the order is preserved for all metals, where values are available (Hg, Cd, Zn,Cu, Mn).

Unconstrained systems:

For CC system, relative order of affinities is preserved in the simplification.

For MM system, the relative order of affinities is preserved, except for Mn, that improves its affinity over Fe and Ni in the simplification.

In HHTCp, the order is preserved for all metals, where values are available (Hg, Cd, Zn, Co, Ni, Mn). In HHTCd, The order is slightly perturbed. The original order of affinities Hg >> Cu >> Cd ~ Co > Zn > Fe > Ni > Mn changes to Hg >> Cu >> Cd ~ Co ~ Zn > Ni ~ Fe >> Mn (>> stands for the difference larger than 7 kcal.mol^{-1} > stands for difference 2-7kcal.mol^{-1} and ~ for difference less than 2 kcal.mol^{-1}). This order is almost identical, except for change of Fe > Ni to Ni ~ Fe.

Although exceptions are present, the relative ordering is mostly preserved. This holds even when the differences are significantly lower than the aforementioned estimated accuracy limit of 10 kcal.mol^{-1}. Should the simplification energies for different metals vary randomly, we could expect more massive perturbations in the relative order of affinities. Although the preservation is aided by "chance" to a certain extent, as preservation of differences smaller than few kilocalories cannot be attributed to the accuracy of DFT method, other contributions seem to take place and the situation calls for more thorough investigation.

We define relative interaction energy E^{rint}_{i} as

$$E_i^{rint} = \frac{\left(E_i^{int} - E_{min}^{int}\right)}{\left(E_{max}^{int} - E_{min}^{int}\right)} \tag{9}$$

where E^{int}_{max} and E^{int}_{min} signify interaction energy of a metal with highest (most negative E^{int}_{i}) and lowest (most positive E^{int}_{i}) affinity in the metal series, respectively. Thus, metal with lowest affinity for given system has $E^{rint}_{i} = 0$, metal with highest affinity has $E^{rint}_{i} = 1$, and metal with $E^{rint}_{i} = 0.5$ satisfies $E^{int}_{i} = (E^{int}_{max}+E^{int}_{min})/2$

Definition of this variable allows us to directly compare complexation energies of original and simplified systems. Table 8 show how does the relative complexation energies change in the simplification ($E^{rint}_{1st,i} - E^{rint}_{orig,1}$).

Table 8: Change of relative interaction energy upon transition from a whole protein system to a simplified system. Both systems were treated with COSMO model.
Values in the table are calculated from relative interaction energy (as defined in (9)) of whole model peptides and their simplified counterparts. Values are dimensionless.

Geom. constraints[a]	system[b]	MN	FE	CO	NI	CU	ZN	CD	HG
free	CC	0.000	0.016	0.047	0.020	0.072	-0.009	-0.017	0.000
	MM	-0.047	-0.008	-0.018	0.019	0.077	-0.002	0.004	0.000
	HHTCp	0.000		0.054	0.024		0.065	0.068	0.000
	HHTCd	0.000	0.004	-0.074	0.100	-0.093	-0.020	-0.071	0.000
fixed	CC	0.000	0.016	0.053	0.052	0.026	0.055	0.014	0.000
	MM	-0.022	0.034	0.021	0.006		-0.030	-0.011	0.000
	HHTCp	0.000				0.088	0.005	-0.013	0.000
	HHTCd	0.000				0.000	-0.085	-0.038	-0.072

[a] … free – no constraints on geometry, fixed – fixed ligand-metal-ligand angles
[b] … whole model peptides (3.2.1) and their simplified counterparts (3.2.2)

The relative order of affinities is preserved as long as these values are small or do not differ significantly within the series. To get an idea what significantly means let us look at the previous estimate. For example, in constrained HHTCd system simplification energies vary from 6.9 to 11.5, i.e. in a small range of 4.6 kcal.mol^{-1}. This means that although the change in relative complexation energies is not negligible (0; -0.038; -0.072; -0.085), it is systematic, i.e. in one direction and of similar magnitude. As a result, relative order of affinities is perfectly preserved.

In unconstrained CC system simplification energies vary in a slightly larger range of 7.8 kcal.mol^{-1}. Change of relative complexation is very small (< 0.05) in all but one case (Cu), and the relative order of affinities is, again, perfectly preserved.

Unconstrained HHTCd system is noteworthy as well. Simplification energies vary in a range of 7.7 kcal.mol^{-1}. This values is more significant than in previous two cases, as it constitutes ca. 15% of the range of complexation energies (range=52.1 kcal.mol^{-1}). However, change of relative complexation energies is smaller and, moreover, similar for subgroups of the series – <-0.093;-0.071> for Co, Cu and Cd and <-0.020;0.004> for Mn, Fe, Hg and Zn. Within these subgroups relative order of affinities is retained. For two metals from different subgroups (i.e. with significantly different change of relative

interaction energy) relative order can generally differ, but thanks to relatively large differences in complexation energies (with respect to relatively small changes of relative interaction energy) it often remains unchanged, as in this case. Ni is an exception with $\Delta E^{\text{rint}} = +0.100$, which is different enough to cause a change position in the relative order of affinities.

In conclusion, described simplification of a protein to a complex containing only transition metal and first-sphere ligands retains basic selectivity properties. Possible improvement of the results might be achieved using more accurate methods, such as MP2, which shows excellent agreement with CCSD(T) calculations, as described in previous chapters. Although MP2 is computationally more challenging, improvements might be worth the cost. Investigating this question will be the subject of future studies.

4.2.2 Gas phase simplified systems

On the other hand, computational cost can be decreased by performing calculations in gas phase environment instead of using COSMO model. This choice can be rationalized by the fact that the inside of a protein is usually hydrophobic and is better approximated by vacuum than by water environment. Feasibility of this decision is investigated in this chapter. Complexation energies of constrained and unconstrained simplified systems are presented in Table 9:

Table 9: Change of relative interaction energy upon transition from a whole protein system to a simplified system. Both systems were treated with COSMO model.
Values in the table are calculated from relative interaction energy (as defined in (9)) of whole model peptides and their simplified counterparts. Values are dimensionless.

Geom. constraints[a]	System[b]	MN	FE	CO	NI	CU	ZN	CD	HG	$E^{int}_{max} - E^{int}_{min}$
Free	CC	-350.0	-346.5	-369.7	-370.9	-382.7	-375.7	-378.4	-436.9	90.4
	MM	77.4	72.4	65.4	66.3	44.4	56.5	42.3	-1.2	78.7
	HHTCp	-202.4	-212.8	-220.4	-218.5	-229.3	-221.4	-221.0	-251.5	49.1
	HHTCd	-394.6	-407.8	-416.3	-415.5	-422.5	-411.6	-402.0	-439.1	44.5
fixed	CC	-347.6	-356.9	-362.4	-365.6	-378.5	-372.9	-377.9	-421.0	73.4
	MM	82.0	80.5	72.9	74.1	55.1	63.1	45.9	-0.3	82.3
	HHTCp	-195.2	-205.8	-214.0	-219.2	-230.5	-210.8	-209.7	-252.2	57.0
	HHTCd	-385.0	-400.9	-408.4	-410.6	-423.1	-396.4	-391.9	-419.7	38.1

[a] … free – no constraints on geometry, fixed – fixed ligand-metal-ligand angles

[b] … simplified systems derived as described in chapter 3.2.2

Charged residues, which are present in all system except MM, are stabilized by water environment resulting in significantly lower complexation energies. However, absence of this stabilization in gas phase environment causes strong Coulomb interaction between these charged residues and transition metal cation. As a result, complexation energies are significantly larger (in absolute value) in gas phase environment, except for MM system, which contains no charged residues.

Nevertheless, the actual range of complexation energies (E^{int}_{max} - E^{int}_{min}) is comparable to the ones found for original systems and simplified systems calculated with COSMO model. However, the range of simplification energies is considerably larger (up to 26.0 kcal.mol^{-1} vs 8.7 kcal.mol^{-1} for unconstrained systems; up to 34.2 vs 8.6 kcal.mol^{-1} for constrained systems).

Previously, it has been argued that wide range of simplification energies does not imply perturbation in relative order of predicted affinities. Examination of relative complexation energies proved to be useful in investigating this problem. We proceed analogically.

Table 10: Change of relative interaction energies for whole peptides calculated with COSMO model and simplified systems calculated in vacuum environment. All values are in kcal.mol^{-1}.

Geom. Constraints[a]	System[b]	MN	FE	CO	NI	CU	ZN	CD	HG
Free	CC	0.039	-0.058	0.173	0.223	-0.058	0.136	-0.070	0.000
	MM	-0.047	0.024	0.051	0.142	-0.083	0.056	-0.040	0.000
	HHTCp	0.000		0.164	0.242		0.105	-0.037	0.000
	HHTCd	0.000	0.088	0.099	0.334	-0.166	0.066	-0.232	0.000
fixed	CC	0.000	0.082	0.133	0.213	0.016	0.202	-0.018	0.000
	MM	-0.022	0.014	0.085	0.096		0.087	-0.043	0.000
	HHTCp	0.000				-0.042	0.105	-0.139	0.000
	HHTCd	0.000				0.000	0.041	-0.143	-0.033

[a] … free – no constraints on geometry, fixed – fixed ligand-metal-ligand angles
[b] … whole model peptides (3.2.1) and their simplified counterparts (3.2.2)

(constrained MM-Cu systems are excluded from discussion for reasons described above)

Inferring from Table 10, the relative complexation energies, change quite drastically. Best correlation of results is found in constrained HHTCd system (ΔE^{rint}_{i} range from -0.143 to 0.041), which is not a particularly good agreement. The worst results are for unconstrained HHTCd system where ΔE^{rint}_{i} range from -0.166 to +0.334! Such a wide range translates into extensive errors in predicted selectivity for metal ions.

Indeed, further examination of relative order of affinities (Figure 8 on the next page) confirms this premonition. Although best and worst binding partners are, mostly, correctly predicted, relative order of affinities of other metals fluctuates chaotically.

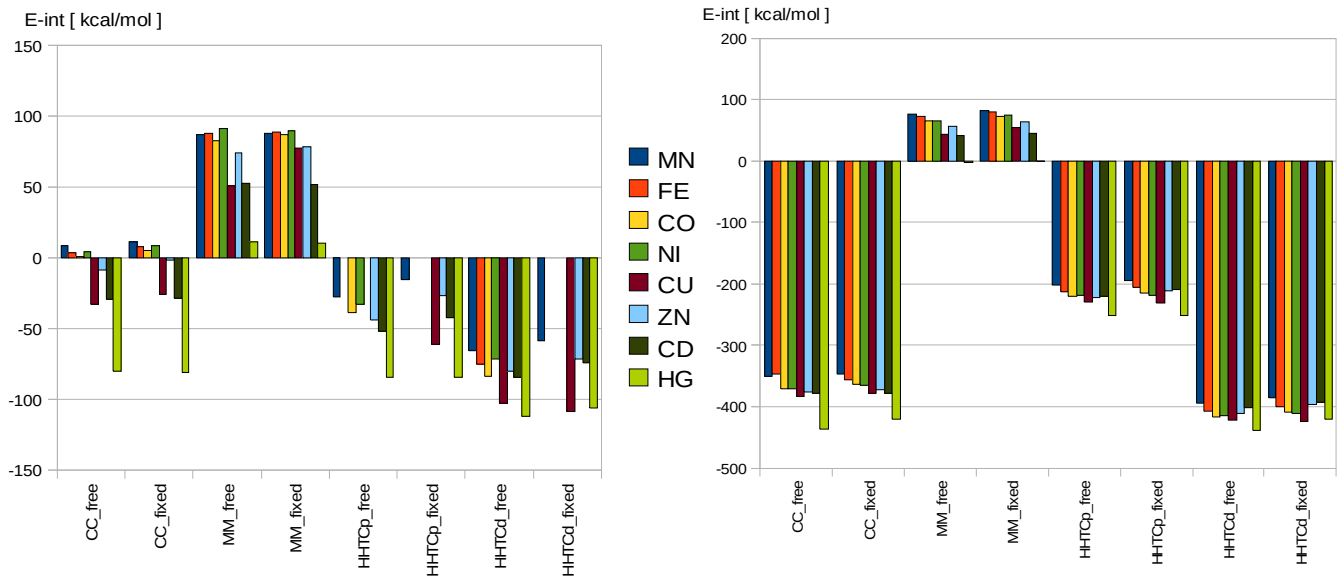


Figure 8. Relative affinities of systems for various metal ions. Interaction energies of eight divalent metal ions with studied systems are displayed.

Left – model peptides, COSMO model

Right – simplified systems, gas phase

CC and MM systems are small tetrapeptides, with metal binding site open to the environment. It is thus not too surprising that approximating these system with gas phase environment fails to preserve the properties of interest. An improvement is expected in the case of icosapeptide HHTCx systems. However, none occurs.

To sum up, representation of a metalloprotein by its transition metal and the first layer of ligands placed in vacuum environment is unsuitable for predicting metal ion selectivity and cannot be used for drawing other than the most basic conclusions about the original system.

4.3 $[MX_n]$ complexes: introduction to combinatorial quantum chemistry

A vast majority of the complexes have been optimized. The *feeder* script allowed almost fully automated treatment of these calculations. Most of the unfinished ones, however, seem to be beyond its capabilities. A more thorough or even individual approach, that will be required to tackle these challenges, is a quest for upcoming future.

Table 11: Converged optimizations of complexes. Status up to August 2010. The number of converged optimizations and the number of remaining ones add up to a total number of different complexes that can be constructed in a manner described in 3.3.1.3

metal	Converged optimizations	remaining	percentage
Mn	15,078	124	99.18%
Fe	14,742	480	96.97%
Co	14,890	312	97.95%
Ni	14,771	341	97.16%
Cu	13,580	1,642	89.33%
Zn	15,155	47	99.69%
Cd	15,059	143	99.06%
Hg	15,123	79	99.48%
total	118,398	3168	97.39%

Optimized geometries will be used for calculation of interaction energies and, thereafter, assessment of metal-selectivity properties. The plan is to perform SPE calculations of the optimized geometries at MP2 level using COSMO model. Based on the results presented and discussed herein, calculation of complexation energies can be used for estimating potential for metal-ion selectivity of studied complexes – and possibly novel peptides derived from these systems – with an accuracy that might not be capable of recognising subtle differences, but should be robust enough for revelation of selectivity trends.

5 Conclusions

Comparison of different methods showed that energies of studied complexes calculated by MP2 are almost identical to CCSD(T). DFT methods does show nonnegligible deviations.

Geometries obtained from optimizations at DFT level show agreement with MP2. Deviations are not uncommon, though. The size of the deviations is generally larger for more complex systems, but can be a result of attaining a different local minimum. The significance of geometry deviation is best assessed by the energy differences it causes. These can be expected to be as high as 15 kcal.mol^{-1}. An idea of DFT-preoptimization is attractive, but further increases computational demands of the project.

Use of COSMO model does influence geometry and energetics of studied systems. The Coulomb interaction has no longer a superior influence on interaction energy of a metal ion with a complex. Interaction energies calculated using COSMO model are significantly closer to realistic values.

A method for simplification of peptide to a complex containing metal ion and a first layer of ligands has been proposed and tested. DFT calculations show that although overall interaction energies change upon simplification of peptide to its model complex, relative order of affinities for different metal ions is well preserved. Exceptions include mostly very small differences in metal affinities that are below the accuracy limit.

COSMO model is necessary for retaining these properties. Analysis of simplified complexes in vacuum environment showed to be unreliable and has been dismissed.

An effort to represent a large number of peptidic metal-binding sites, that would include different geometries, metal ions and binding residues, has been undertaken. The approach includes construction of all combinations that vary eleven ligands representing amino acid side chain residues, five geometries and eight metal ions. Initial guess of geometries was obtained for all these complexes and these geometries were optimized at DFT level. Highly automated and sophisticated treatment of vast number of systems allowed most of the optimizations to be successfully achieved, while the rest are to be dealt with in the upcoming future.

Results of method verification suggest the obtained data could be used directly for calculation of interaction energies at MP2 level, or further optimized at MP2 level if higher level of accuracy is demanded.

6 Literature

1. Bertini, I.; Gray, H. B.; Stiefel, E. I.; Valentine J. S. *Biological inorganic chemistry: structure and reactivity, University Science Books, Sausalito, Calif, 2007*
2. Thompson, K.H.; Orvig, C. *Science*, **2003**, *300, 936-939*
3. Ragsdale, S.W. *Chem. Rev.* **2006**, *106*, 3317-3337
4. Andreini, C.; Bertini, I.; Rosato, A. *Bioinformatics* **2004**,*20*, 1373-1380.
5. Kamerlin, S. C. L.; Warshel, A. *Proteins* **2010**, *78*, 1339-1375.
6. Senn, H. M.; Thiel, W. *Angew. Chem. Int. Edit.* **2009**, *48*, 1198-1229.
7. Holm, R. H.; Kennepohl, P.; Solomon E. I. *Chem. Rev.* **1996**, *96*, 2239-2314.
8. Jakab, I. N.; Jenei, Z.; Gyurcsik, B.; Gajda, T.; Körtvélyesi, T.; Mikulová, A.; Rulíšek, L.; Raskó, T.; Kiss, A. in A*chievements in Coordination, Bioinorganic and Applied Inorganic Chemistry*, Melník, M.; Šima, J.; Tatarko, M.; Eds.; Slovak Technical University Press: Bratislava, 2007, pp. 80–90.)
9. Rulíšek, L.; Vondrášek, J. *J. Inorg. Biochem.* **1998**, *71*, 115-127.
10. Dupont, C. L.; Yang, S.; Palenik, B.; Bourne, P. E.; Proc. *Natl Acad. Sci. U. S. A.*, **2006,** *103, 17822-17827*
11. Lander, E.S.; Linton, L.M.; Birren, B.; Nusbaum, C.; Zody, M.C.; Baldwin, J.; Devon, K.; Dewar, K.; Doyle, M.; FitzHugh, W.; Funke, R.; Gage, D.; Harris, K.; Heaford, A.; Howland, J.; Kann, L.; Lehoczky, J.; LeVine, R.; McEwan, P.; McKernan, K.; Meldrim, J.; Mesirov, J.P.; Miranda, C.; Morris, W.; Naylor, J.; Raymond, C.; Rosetti, M.; Santos, R.; Sheridan, A.; Sougnez, C.; Stange-Thomann, N.; Stojanovic, N.; Subramanian, A.; Wyman, D.; Rogers, J.; Sulston, J.; Ainscough, R.; Beck, S.; Bentley, D.; Burton, J.; Clee, C.; Carter, N.; Coulson, A.; Deadman, R.; Deloukas, P.; Dunham, A.; Dunham, I.; Durbin, R.; French, L.; Grafham, D.; Gregory, S.; Hubbard, T.; Humphray, S.; Hunt, A.; Jones, M.; Lloyd, C.; McMurray, A.; Matthews, L.; Mercer, S.; Milne, S.; Mullikin, J.C.; Mungall, A.; Plumb, R.; Ross, M.; Shownkeen, R.; Sims, S.; Waterston, R.H.; Wilson, R.K.; Hillier, L.W.; McPherson, J.D.; Marra, M.A.; Mardis, E.R.; Fulton, L.A.; Chinwalla, A.T.; Pepin, K.H.; Gish, W.R.; Chissoe, S.L.; Wendl, M.C.; Delehaunty, K.D.; Miner, T.L.;

Delehaunty, A.; Kramer, J.B.; Cook, L.L.; Fulton, R.S.; Johnson, D.L.; Minx, P.J.; Clifton, S.W.; Hawkins, T.; Branscomb, E.; Predki, P.; Richardson, P.; Wenning, S.; Slezak, T.; Doggett, N.; Cheng, J.F.; Olsen, A.; Lucas, S.; Elkin, C.; Uberbacher, E.; Frazier, M.; Gibbs, R.A.; Muzny, D.M.; Scherer, S.E.; Bouck, J.B.; Sodergren, E.J.; Worley, K.C.; Rives, C.M.; Gorrell, J.H.; Metzker, M.L.; Naylor, S.L.; Kucherlapati, R.S.; Nelson, D.L.; Weinstock, G.M.; Sakaki, Y.; Fujiyama, A.; Hattori, M.; Yada, T.; Toyoda, A.; Itoh, T.; Kawagoe, C.; Watanabe, H.; Totoki, Y.; Taylor, T.; Weissenbach, J.; Heilig, R.; Saurin, W.; Artiguenave, F.; Brottier, P.; Bruls, T.; Pelletier, E.; Robert, C.; Wincker, P.; Smith, D.R.; Doucette-Stamm, L.; Rubenfield, M.; Weinstock, K.; Lee, H.M.; Dubois, J.; Rosenthal, A.; Platzer, M.; Nyakatura, G.; Taudien, S.; Rump, A.; Yang, H.; Yu, J.; Wang, J.; Huang, G.; Gu, J.; Hood, L.; Rowen, L.; Madan, A.; Qin, S.; Davis, R.W.; Federspiel, N.A.; Abola, A.P.; Proctor, M.J.; Myers, R.M.; Schmutz, J.; Dickson, M.; Grimwood, J.; Cox, D.R.; Olson, M.V.; Kaul, R.; Raymond, C.; Shimizu, N.; Kawasaki, K.; Minoshima, S.; Evans, G.A.; Athanasiou, M.; Schultz, R.; Roe, B.A.; Chen, F.; Pan, H.; Ramser, J.; Lehrach, H.; Reinhardt, R.; McCombie, W.R.; de, la, B.astide, M.; Dedhia, N.; BlÃ¶cker, H.; Hornischer, K.; Nordsiek, G.; Agarwala, R.; Aravind, L.; Bailey, J.A.; Bateman, A.; Batzoglou, S.; Birney, E.; Bork, P.; Brown, D.G.; Burge, C.B.; Cerutti, L.; Chen, H.C.; Church, D.; Clamp, M.; Copley, R.R.; Doerks, T.; Eddy, S.R.; Eichler, E.E.; Furey, T.S.; Galagan, J.; Gilbert, J.G.; Harmon, C.; Hayashizaki, Y.; Haussler, D.; Hermjakob, H.; Hokamp, K.; Jang, W.; Johnson, L.S.; Jones, T.A.; Kasif, S.; Kaspryzk, A.; Kennedy, S.; Kent, W.J.; Kitts, P.; Koonin, E.V.; Korf, I.; Kulp, D.; Lancet, D.; Lowe, T.M.; McLysaght, A.; Mikkelsen, T.; Moran, J.V.; Mulder, N.; Pollara, V.J.; Ponting, C.P.; Schuler, G.; Schultz, J.; Slater, G.; Smit, A.F.; Stupka, E.; Szustakowski, J.; Thierry-Mieg, D.; Thierry-Mieg, J.; Wagner, L.; Wallis, J.; Wheeler, R.; Williams, A.; Wolf, Y.I.; Wolfe, K.H.; Yang, S.P.; Yeh, R.F.; Collins, F.; Guyer, M.S.; Peterson, J.; Felsenfeld, A.; Wetterstrand, K.A.; Patrinos, A.; Morgan, M.J.; de, J.ong, P.; Catanese, J.J.; Osoegawa, K.; Shizuya, H.; Choi, S.; Chen, Y.J.; International, Human, Genome, Sequencing, Consortium. *Nature*, **2001**, *409,860-921*

12. Maret, W. *Antioxid. Redox Sign.* **2006**, *8, 1419-1441*

13. Andreini, C; Bertini, I.; Cavallaro, G; Hollidat, G. L.; Thornron, J. M. *J. Biol. Inorg. Chem.,* **2008**, *13,* 1205-1218
14. Roach, P.L.; Clifton, I.J.; Hensgens, C.M.; Shibata, N.; Schofield, C.J.; Hajdu, J.; Baldwin, J.E. *Nature,* **1997**, *387,827-830*
15. Andreini, C; Bertini, I.; Cavallaro, G; Hollidat, G. L.; Thornron, J. M. *J. Biol. Inorg. Chem.,* **2008**, *13, 1205-1218*
16. Blaszczyk, J.; Shi, G.; Yan, H.; Ji, X. *Structure,* **2000**, *8,* 1049-1058
17. Li, Y.; Blaszczyk, J.; Wu, Y.; Shi, G.; Ji, X.; Yan, H. *Biochemistry,* **2005**, *44, 8590-8599*
18. Carpenter, E.P.; Hawkins, A.R.; Frost, J.W.; Brown, K.A.; *Nature,* **1998**, *394, 299-302*
19. Essen, L.O.; Perisic, O.; Katan, M.; Wu, Y.; Roberts, M.F.; Williams, R.L. *Biochemistry,* **1997**, *36, 1704-1718*
20 . Solomon, E.I.; Sundaram, U.M.; Machonkin, T.E. *Chem. Rev.* **1996**, *96,* 2563-2605.
21. Rosenzweig, A. C. *Acc. Chem. Res.*, **2001**, *34, 119-128*
22. Culotta, V. C.; Yang, M; O'Halloran, T. V. *Biochim. Biophys. Acta,* **2006**, *1763*, 747–758
23. Labbé, R. F.; Dewanji, A. *Clin. Biochem.*, **2004**, *37,* 165–174
24. O'Halloran, T. V. *Science***,1993**, *261***,** *715–725*
25. Giedroc, D. P.; Arunkumar, A. I. *Dalton Trans.*,**2007**, *3107–3120*
26. Waldron, K. J.; Robinson, N. J. *Nature Rev. Microbiol.***,2009**, *7, 25–35*
27. Changela, A.; Chen, K.; Xue, Y.; Holschen, J.; Outten, C. E.,; O'Halloran, T. V.; Mondragón, A. *Science*, **2003**,*5, 1383-1387.*
28. Cavet, J.S.; Meng, W; Pennella, M.A.; Appelhoff, R.J.; Giedroc, D.P; Robinson, N.J. *J. Biol. Chem.,* **2002**,*277, 38441–38448*
29. Guedon, E.; Helmann, J. D. *Mol. Microbiol.* ,**2003**, *48,* 495–506

30. Golynskiy, M. V.; Gunderson, W. A.; Hendrich, M. P.; Cohen, S. M. *Biochemistry,* **2006**, *45, 15359–15372*

31. Phillips, C.M.; Schreiter, E.R.; Guo, Y.; Wang, S.C.; Zamble, D.B.; Drennan, C.L. *Biochemistry*,**2008**,*47,* 1938–1946

32. Vasak, M.; Hasler, D. W. *Curr. Opin. Chem. Biol.,* **2000**, *4,* 177

33. Ge, R.; Watt, R. M.; Sun, X.; Tanner, J. A.; He, Q. Y.; Huang, J. D.; Sun, H. *Biochem. J.*,**2006**, *393, 285.*

34. Gilbert, J. V.; Ramakrishna, J.; Sunderman, F. W.Jr.; Wright, A.; Plaut, A. G. *Infect.Immun.,* **1995**, *63, 2682.*

35. Cotton, F. A.; Wilkinson, G. *Advanced Inorganic Chemistry, Wiley & Sons, Inc., New York, **1999***

36. Parr, R. G.; Pearson, R. G. *J. Am. Chem. Soc.* **1983**, *105*, 7512-7516

37. Warshel, A. *Computer Modeling of Chemical Reactions in Enzymes and Solutions. John Wiley&Sons, Inc., New York, 1997.*

38 Fraústo da Silva, J.J.R.; Williams, R.J.P. *The Biological Chemistry of the Elements. 2nd edition, Oxford University Press, Inc.: New York, **2001**.*

39. Berthon, G. *Handbook of Metal–Ligand Interactions in Biological Fluids, Ed.; Vol. 2, Marcel Dekker: New York, **1995**.*

40. Sigel, R.K.O.; Sigel, H. *Acc. Chem. Res.* **2010**, *43*, 974-984.

41. Dudev, T.; Lim, C. *Chem. Rev.* **2003**, *103*, 773–787.

42. Williams, R. J. P. *Biometals* **2007**, *20*, 107-112.

43. Riccardi, D.; Shaefer, P.; Cui, Q. *J. Phys. Chem. B* **2005**, *109*, 17715–17733.

44. Rulíšek, L.; Havlas, Z. *J. Chem. Phys.* **2000**, *112*, 149-157.

45. Rulíšek, L.; Šponer, J. *J. Phys. Chem. B* **2003**, *107*, 1913-1923.

46. Wladkowski, B. D.; Krauss, M.; Marti, M. A.; Stevens, W. J. *J.Phys. Chem.* **1995**, *99*, 6273–6276.

47. Shurki, A.; Warshel, A. *Proteins* **2004**, *55*, 1–10.

48. Shurki, A.; Warshel, A. *Adv. Protein Chem.* **2003**, *66*, 249–312.

49. Kamerlin, S.C.L.; Haranczyk, M.; Warshel, A. *J. Phys. Chem. B* **2009**, *113*, *1253-1272.*

50. Rulíšek, L.; Havlas, Z. *J. Am. Chem. Soc.* **2000**, *122*, 10428-10439.

51. Rulíšek, L.; Havlas, Z. *J. Phys. Chem. A* **2002**, *106*, 3855-3866.

52. Rulíšek, L.; Havlas, Z. *Int. J. Quantum Chem.* **2003**, *91*, 504-510.

53. Rulíšek, L.; Havlas, Z. *J. Phys. Chem. B* **2003**, *107*, 2376-2385.

54. Dudev, T.; Lim, C. *J. Phys. Chem. B* **2001**, *105*, 10709–10714.

55. Dudev, T.; Lim, C. *Acc. Chem. Res.* **2007**, *40*, 85–93.

56. Blumberger, J.; Sprik, M. *Theor. Chem. Acc.* **2006**, *115*, 113–126.

57. Haranczyk, M.; Gutowski, M.; Warshel, A. *Phys. Chem. Chem. Phys.* **2008**, *10*, 4442–4448.

58. Simonson, T.; Carlsson, J.; Case, D. *J. Am. Chem. Soc.* **2004**, *126*, 4167–4180.

59. Lu, Y.; Yeung, N.; Sieracki, N.; Marshall, N. M. *Nature*, **2009**,*460*, 855-862

60. DeGrado, W. F.; Summa, C. M.; Pavone, V.; Nastri, F.; Lombardi, A. *Annu. Rev. Biochem.*,**1999**, *68*, 779-819

61. Koder, R. L.; Anderson, J.L.R., Solomon, L.A.; Reddy, K.S.; Moser, C.C.; Dutton, P.L. *Nature*, **2009**, *458*, 305-309

62. Lu, Y.; Berry, S. M.; Pfister, T. D. *Chem. Rev.*,**2001**, *101*, 3047-3080

63. T. Hayashi, Y. Hisaeda, *Acc. Chem. Res.* **2002**, *35*, 35–43.
64. D. W. Low, M. G. Hill, *J. Am. Chem. Soc.* **2000**, *122*, 11039–11040.
65. S. M. Berry, M. D. Gieselman, M. J. Nilges, W. A. Van der Donk, Y. Lu, *J. Am. Chem. Soc.* **2002**, *124*, 2084–2085.
66. T. D. Pfister, A. Y. Mirarefi, A. J. Gengenbach, X. Zhao, C. Danstrom, N. Conatser, Y. G. Gao, H. Robinson, C. F. Zukoski, A. H. J. Wang, Y. Lu, *J. Biol. Inorg. Chem.* **2007**, *12*, 126–137.
67. J. Kaplan, W. F. DeGrado, *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 11566–11570.
68. M. Allert, M. A. Dwyer, H. W. Hellinga, *J. Mol. Biol.* **2007**, *366*, 945–953.
69. Y. Lu, J. S. Valentine, *Curr. Opin. Struct. Biol.* **1997**, *7*, 495–500.
70. D. E. Benson, M. S. Wisz, H. W. Hellinga, *Curr. Opin. Biotechnol.* **1998**, *9*, 370–376.
71. W. F. DeGrado, C. M. Summa, V. Pavone, F. Nastri, A. Lombardi, *Annu. Rev. Biochem.* **1999**, *68*, 779–819.
72. D. Jantz, B. T. Amann, G. J. Gatto, Jr, J. M. Berg, *Chem. Rev.* **2004,** *104*, 789–799.
73. C. J. Reedy, B. R. Gibney, *Chem. Rev.* **2004**, *104*, 617–649.
74. S. S. Huang, B. R. Gibney, S. Stayrook, P. L. Dutton, M. Lewis, *J. Mol. Biol.* **2003**, *326*, 1219–1225.
75. B. T. Farrer, N. P. Harris, K. E. Balchus, V. L. Pecoraro, *Biochemistry* **2001**, *40*, 14696–14705.
76. B. Farrer, C. McClure, J. E. Penner-Hahn, V. L. Pecoraro, *Inorg. Chem.* **2000**, *39*, 5422-5423.
77. M. Matzapetakis, B. T. Farrer, T.-C. Weng, L. Hemmingsen, J. E. Penner-Hahn, V. L. Pecoraro, *J. Am. Chem. Soc.* **2002**, *124*, 8042–8054.
78. A. K. Petros, A. R. Reddi, M. L. Kennedy, A. G. Hyslop, B. R. Gibney, *Inorg. Chem.* **2006**, *45*, 9941–9958.
79. M. Matzapetakis, D. Ghosh, T. C. Weng, J. E. Penner-Hahn, V. L. Pecoraro, *J. Biol. Inorg. Chem.* **2006**, *11*, 876–890.
80. Lu, Y. *Curr. Opin. Chem. Biol.* **2005**, *9*, 118-126
81. Ikeda, Y.; Kawahara, S.; Taki, M.; Kuno, A.; Hasegawa, T.; Taira, K. *Protein Eng.*, **2003**, *16*, 699-706
82. Noren, C. J.; Anthony-Cahill, S. J.; Griffith, M. C.; Schultz, P. G. *Science***, 1989**, *244*,

182-188

83. Low, D. W.; Hill, M. G. *J. Am. Chem. Soc.*, ***2000***,*122*, 11039-1104

84. Wilson, M. E.; Whitesides, G. M.; *J. Am. Chem. Soc.*, **1978**, *100*, 306-307

85. Collot, J; Humbert, N.; Skander, M.; Klein, G.; Ward, T.R *J. Am. Chem. Soc.*, **2003**, *125*, 9030-9031

86. Steinreiber, J.; Ward, T. R. *Coord. Chem. Rev.*, ***2008,*** *252*, 751-766

87. Letondor, C.; Pordea, A.; Humbert, N.; Ivanova, A.; Mazurek, S.; Novic, M.; Ward, T.R. *J. Am. Chem. Soc.***, 2006**, *128*, 8320-8328

88. Creus, M.; Pordea, A.; Roseel, T.; Sardo, A.; Letondor, C.; Ivanova, A.; LeTrong, I.; Stenkamp, R. E.; Ward, T.R. *Angew. Chem. Int. Edit. Engl*., **2008**, *47*, 1400-1404

89. Kotrba, P.; Pospíšil, P.; Lorenzo, V.; Ruml, T. *J. Recept. Signal Transduct. Res.* **1999**, *19*, 703–715.

90. Kotrba, P.; Dolečková, L.; Lorenzo, V.; Ruml, T., *Appl. Environ. Microbiol.* **1999**, *65*, 1092–1098.

91. Sousa,. C; Kotrba, P.; Ruml, T.; Cebolla, A.; Lorenzo, V.;, *J. Bacteriol.* **1998**, *180*, 2280–2284.

92. Sousa,. C; Cebolla, A.; Lorenzo, V., *Nat. Biotechnol*. **1996**, *14*, 1017–1020.

93. Macek, T.; Kotrba, P.; Svatoš, A.; Novákova, M.; Demnerova, K.; Mackova, M. *Trends Biotechnol.* **2008**, *26*, 146-152.

94. S. Brown *Nat. Biotechnol.* **1997**, *15*, 269–272.

95. Dahiyat B. I.; Mayo, S. L. *Science* **1997**, *278*, 82–87.

96. Kostal, J.; Mulchandani, A.; Chen, W. *Macromolecules* **2001**, *34*, 2257–2261.

97. Ramanathan, S.; Ensor, M.; Daunert, S. *Trends Biotechnol.* **1997**, *15*, 500–506.

98. Marvin, J. S.; Hellinga, H.W. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 4955–4960.

99. Bontidean, I.; Kumar, A.; Csoregi, E.; Galaev I. Y.; Mattiasson, B. *Angew Chem. Int. Edit. Engl.* **2001**, *40*, 2676–2678.

100. Gibney, B. R.; Rabanal, F.; Skalicky, J.J.; Wand, A.J.; Dutton, P.L. *J. Am. Chem. Soc.* **1999**, *121*, 4952–4960.

101. Chen, X.; Discher, B.M.; Pilloud, D.L.; Gibney, B.R.; Moser, C.C.; Dutton P.L. *J.*

Phys. Chem. B **2002**, *106*, 617–624.

102. Szabo, A.; Ostlund, S.N. *Modern quantum chemistry*. McGraw-Hill, Inc. , New York, **1989**

103. Jensen, F. *Introduction to computational chemistry*. John Wiley & Sons, , Inc., New York, **2007**

104. Cramer, J. C. *Essentials of computational chemistry: theories and models. John Wiley&Sons, Inc., New York,* **2004**.

105. Born, M.; Oppenheimer, J.R. *Ann. Phys. (Leipzig)*, **1927**, 84, *457*

106. Hirata, S.; Miller, E.B. *J. Phys. Chem. A,* **2009**, *113, 12461-12469*

107. Zettili, N.; Villars, F. M. H. *Nuclear Physics A*, **1987**, *469, 93-105*

108. Kim, K.; Jordan, D. *J. Phys. Chem.*, **1994**, *98*, 10089-10094

109. Alfredsson, M.; Ojamaė, L.; Hermansson, K. G. *Int. J. Quantum Chem.*, **1998**, *60*, 767-777

110. Watts, J. D.; Gauss, J; Bartlett, R. J. *J. Chem. Phys.*, **1993**, *98*, 8718

111. Raghavachari, K.; Trycks, G. W.; Pople J. A.; Head-Gordon, M. *Chem. Phys. Lett. ,* **1989**, *157, 479*

112. Parr, R.; Yang, W. *Density functional theory of atoms and molecules, Oxfor University Press, New York, **1989***

113. Kohn, W.; Sham, L. J. *Phys. Rev. A* **1965**, *140*, 1133.

114. Staroverov, V. N.; Scuseria, G. E.; Tao, J.; Perdew, J. P. *J. Chem. Phys.* **2003**, *119*, 12129–12137.

115. (a) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098-3100. (b) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785-789. (c) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648-5652. (d) Stephens, P. J.; Devlin, F. J.; Frisch, M. J.; Chabalowski, C. F. *J. Phys. Chem.* **1994**, *98*, 11623-11627.

116. Hertwig, R. H.; Koch W. *Chem. Phys. Lett.* **1997**, *268*, 345-351.

117. Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.*, **1996**, *77*, 3865-3868.

118. Deglmann, P.; Furche, F.; Ahlrichs, R. *Chem. Phys. Lett.,* **2002**, *5-6, 511-518*

119. Hobza, P.; Müller-Dethlefs, K. *Non-covalent interactions: Theory and experiment, The Royal Society of Chemistry, London,* **2010**

120. Korth, M.; Grimme, S. *J. Chem. Theory Comput.,* **2009**, *5,* 993-1003

121. Boys, S. F.; Bemardi, F. *Mol. Phys,* **1970***, 19, 553*

122. Klamt, A.; Schüürmann, G. *J. Chem. Soc., Perkin Trans. ,* **1993***, 2, 799-805*

123. Takano, Y.; Houk, K. N. *J. Chem. Theory Comput. 1*, **2005**, *70-77*

124. (a) Ahlrichs, R.; Bär, M.; Häser, M.; Horn, H.; Kolmel, C. *Chem. Phys. Lett.*, **1989***,162,*165. (b) Häser, M.; Weigend, F. *Theor. Chem. Acc.,* **1997**, *97,* 331. (c) Hättig, C.; Weigend, F., *J. Chem. Phys*, **2000**, *113, 5154.* (d) Weigend, F. *Phys. Chem. Chem. Phys.,* **2004**, *4,* 4285

125. Eichkorn, K.; Treutler, O.; Öhm, H.; Häser, M.; Ahlrichs, R. *Chem. Phys. Lett.* **1995**, *240*, 283-290.

126. Eichkorn, K.; Weigen, F.; Treutler, O.; Ahlrichs, R. *Theor. Chim. Acta* **1997***, 97*, 119-124.

127. Weigend, F.; Ahlrichs, R. *Phys. Chem. Chem. Phys.*, **2005** *7*, 3297-3305.

128. Schäfer, A.; Huber, C.; Ahlrichs, R. *J. Chem. Phys.* **1994**, *100*, 5829-5835.

129. Woon, D. E.; Dunning, T. H., Jr. *J. Chem. Phys.* **1993**, *98*, 1358-1371.

130. Kendall, R. A.; Dunning, T. H., Jr.; Harrison, R. J. *J. Chem. Phys.* **1992**, *96*, 6796-6806.

131. Klamt, A.; Schuurmann, G. *J. Chem. Soc.-Perkin Trans. 2* **1993**, 799-805.

132. Schäfer, A.; Klamt, A.; Sattel, D.; Lohrenz, J. C. W.; Eckert, F. *Phys. Chem. Chem. Phys.* **2000,** *2*, 2187-2193.

133. Kožíšek, M.; Svatoš, A.; Buděšínský, M.; Muck, A.; Bauer, M. C.; Kotrba, P.; Ruml, T.; Havlas, Z.; Linse, S.; Rulíšek, L. *Chem. Eur. J.* **2008**, *14*, 7836-7846.