

Modeling Metal Binding Sites in Proteins by Quantum Chemical Calculations

Todor Dudev

Faculty of Chemistry and Pharmacy, University of Sofia, Sofia, Bulgaria
Email: t.dudev@chem.uni-sofia.bg

Received 10 February 2014; revised 12 March 2014; accepted 20 March 2014

Copyright © 2014 by author and Scientific Research Publishing Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Currently about half of all known proteins contain metal cofactors. Metal cations are indispensable components of the cellular machinery and are involved in a number of essential tasks ranging from nucleic acid and protein structure stabilization to enzyme catalysis, signal transduction, muscle contraction, hormone secretion, taste and pain sensation, respiration, and photosynthesis [1]-[3]. They are the simplest, but most versatile, cofactors in protein biochemistry with a plethora of distinctive properties, such as electron-acceptor ability, positive charge, flexible coordination sphere, specific ligand affinity, varying valence state, low- or high-spin configuration, and mobility.

Quantum chemical calculations have been extensively used to model active sites in metalloproteins and evaluate thermodynamic and/or kinetic parameters of the respective biochemical processes. These calculations are well suited for the task as they can capture all the electronic effects (including polarization and charge transfer effects between interacting entities) which are of paramount importance for properly describing the interactions between metal cations and protein ligands. Note that the polarization and charge transfer effects, which are poorly treated (if at all) in other computational approaches based on classical force fields, are quite significant and need to be explicitly accounted for especially in binding sites containing doubly (e.g. Mg^{2+} or Zn^{2+}) or triply charged (Fe^{3+}) metal cations.

In modeling studies usually the metal and its first and/or second ligation sphere are treated by high-level quantum chemical calculations, whereas the rest of the protein is subjected to either continuum dielectric calculations, molecular mechanics computations or lower basis set ab initio/DFT calculations. What is the succession and type of steps that should be undertaken in order to build up a realistic model of the protein binding site and achieve sound computational results? Firstly, the metal and its partner ligands should be properly modeled. For metals, it is quite important to correctly model their coordination number and symmetry of the respective protein or aqua complexes, and spin-state (for open-shell metals) as these factors, along with the metal charge, are among the major determinants of metal binding and selectivity in protein systems [4]. In the model ligand molecule, the chemical nature of the metal-coordinating protein residue should be preserved. Thus, acetate (CH_3COO^-) is a good model for the Asp side chain, while both acetate and propionate (CH_3COO^- and $CH_3CH_2COO^-$) are appropriate models for the Glu side chain. On the other hand, however, formaldehyde (H_2CO) or acetone (CH_3COCH_3), although both containing a carbonyl group, are not proper models for the backbone peptide group ($-CO-NH-$) or side chain of Asn and Gln ($-CO-NH_2$) since the chemical and electronic properties of the amide group differ quite significantly from those of the aldehyde and ketone moieties, respec-

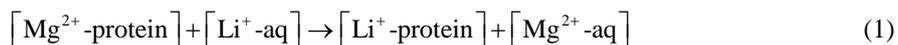
tively. Along the same vein, ammonia (NH₃) or aliphatic amine (CH₃NH₂) are not right models for the ligating –N= center of histidine, as the electron density within the His imidazole ring, due to conjugation effects, is quite different from that in ammonia and aliphatic amines.

Next step is the choice of appropriate level of theory for executing the calculations. Generally, this will depend on the size (number of atoms) of the protein system to be studied and computer resources available. Among different theoretical methods and basis sets which appear affordable for the user, however, the combination which best reproduces a set of experimental observables related to the protein system under study, should be chosen. There is a variety of experimental data suitable for calibrating the ab initio/DFT calculations, such as geometrical parameters (bond distances and valence angles) of metal complexes with water or simple organic/inorganic ligands, gas-phase dipole moments and polarizabilities of protein ligands or their models, enthalpies/free energies of protonation or deprotonation reactions in the gas-phase of given set of ligands, etc.

Biochemical reactions in proteins take place in condensed phase where the active site is located in a cavity/crevice characterized by particular dielectric constant (varying from ~4 for buried binding sites to ~30 for solvent exposed binding pockets). Once the proper level of quantum chemical calculations for treating the metal and its immediate surroundings had been established (see above), the next step should be undertaken: the calibration of the condensed phase calculations. Experimental free energies of solvation in different solvents can be used for the calibration purposes. For performing continuum dielectric calculations, molecular parameters, such as effective atomic van der Waals radii or atomic charges, can be optimized in order to evaluate free energies of solvation that best reproduce the experimental observables.

Before proceeding to the actual calculations (“production run”), a rigorous test—the validation of the calculations—should be done. Experimentally well-characterized reaction in condensed phase which mimics the biochemical process of interest can be chosen and used for that purpose. Successful reproduction of the experimental quantities (within 1 - 2 kcal/mol for the energy/free energy, for example) is an indicator for the calculations’ reliability and robustness.

Modeling the process of Li⁺ → Mg²⁺ substitution in magnesium binding sites in proteins will be used to illustrate the procedure outlined above. The competition between Mg²⁺ and Li⁺ for the metal-binding site can be assessed by the free energy of the Li⁺ → Mg²⁺ exchange reaction:



In Equation (1), [M^{q+}-protein] and [M^{q+}-aq] represent the metal ion (M^{q+} = Mg²⁺ or Li⁺) bound to protein ligands and unbound in the vicinity of the binding cavity, respectively. The metal and its first ligation sphere are treated by quantum chemical methods, while the rest of the protein is modeled as continuum dielectric characterized by a given dielectric constant.

In protein binding sites Mg²⁺ is usually bound to the side chains of Asp/Glu or Asn/Gln, or the backbone peptide group [5]. Accordingly, the side chains of Asp/Glu, Asn/Gln, and peptide backbone group are modeled as acetate (CH₃COO⁻), acetamide (CH₃CONH₂), and N-methylacetamide (CH₃CONHCH₃), respectively. In proteins as in aqueous solution Mg²⁺ is mostly hexacoordinated [6] [7]. Thus, Mg²⁺ complexes are modeled as MgL₆ (L = H₂O, CH₃CONH₂, CH₃CONHCH₃, CH₃COO⁻). The most common coordination number of Li⁺ is four [8]. Accordingly, its complexes are modeled as LiL₄.

The B3LYP/6-31+G(3d, p) method/basis set combination is chosen among various other combinations as it best reproduces the experimental dipole moments, μ, of model ligands and metal—ligand bond distances, R, in metal complexes [9] [10]:

$$\begin{aligned} \mu_{\text{calc}}(\text{H}_2\text{O}) &= 1.88\text{D}, \quad \mu_{\text{exp}}(\text{H}_2\text{O}) = 1.85 \pm 0.02\text{D} \\ \mu_{\text{calc}}(\text{HCONH}_2) &= 3.99\text{D}, \quad \mu_{\text{exp}}(\text{HCONH}_2) = 3.73 \pm 0.07\text{D} \\ \text{in } [\text{Li}(\text{H}_2\text{O})_4]^+ : R(\text{Li-O})_{\text{calc}} &= 1.95 \text{ \AA}, \quad R(\text{Li-O})_{\text{exp}} = 1.94 \pm 0.05 \text{ \AA}, \\ \text{in } [\text{Mg}(\text{H}_2\text{O})_6]^{2+} : R(\text{Mg-O})_{\text{calc}} &= 2.10 \text{ \AA}, \quad R(\text{Mg-O})_{\text{exp}} = 2.07 \pm 0.03 \text{ \AA}. \end{aligned}$$

Condensed-phase calculations are calibrated with respect to experimental free energies of hydration of the metal cations and model ligands. Results are given in **Table 1**. For validation purposes, the experimental free energies of Li⁺ → Mg²⁺ substitution in acetate, oxalate and nitrilotriacetic acid complexes are used and com-

Table 1. Comparison between computed and experimental hydration free energies, $\Delta G_{\text{sol}}^{80}$, of metal cations and model ligands (in kcal/mol) [9].

Molecule	$\Delta G_{\text{sol}}^{80}$ (kcal/mol)		
	Exp.	Calc.	Error ^a
Mg ²⁺	-455.5	-456.3	-0.8
Li ⁺	-123.5	-123.6	-0.1
CH ₃ COO ⁻	-82.2	-82.3	-0.1
CH ₃ CONHCH ₃	-9.7	-9.8	-0.1

^aError = $\Delta G_{\text{sol}}^{80}$ (Calc) - $\Delta G_{\text{sol}}^{80}$ (Exp).

Table 2. Comparison Between Computed and Experimental Free Energies (in kcal/mol) for [MgX] + [Li(H₂O)₄]⁺ → [LiY] + [Mg(H₂O)₆]²⁺ in Water, ΔG^{80} (Li⁺ → Mg²⁺) [9].

[MgX]	[LiY]	ΔG^{80} (Li ⁺ → Mg ²⁺)		
		Exp.	Calc.	Error ^a
[Mg(H ₂ O) ₅ (CH ₃ COO ⁻) ⁺	[Li(H ₂ O) ₃ (CH ₃ COO ⁻) ⁰	0.8	1.3	0.5
[Mg(H ₂ O) ₄ (C ₂ O ₄ ²⁻) ⁰	[Li(H ₂ O) ₂ (C ₂ O ₄ ⁻) ⁻	2.7	2.1	-0.6
[Mg(H ₂ O) ₂ (NTA ³⁻) ⁻ b	[Li(NTA ³⁻) ²⁻	4.1	4.1	0.0

^aError = ΔG^{80} (Calc) - ΔG^{80} (Exp). ^bNTA = nitrilotriacetic acid bound in a tetradentate fashion (including central N atom) to the metal.

pared with those evaluated theoretically using the models and methods described above. As seen from the results presented in **Table 2**, the calculations reproduce the experimentally determined free energies of Li⁺ → Mg²⁺ substitution within 1 kcal/mol. Now, avenues are open for the “production run”.

References

- [1] Frausto da Silva, J.J.R. and Williams, R.J.P. (1991) *The Biological Chemistry of the Elements*. Oxford University Press, Oxford.
- [2] Lippard, S.J. and Berg, J.M. (1994) *Principles of Bioinorganic Chemistry*. University Science Books, Mill Valley.
- [3] Bertini, I., Sigel, A. and Sigel, H. (2001) *Handbook on Metalloproteins*. Marcel Dekker, New York.
- [4] Dudev, T. and Lim, C. (2014) Competition among Metal Ions for Protein Binding Sites: Determinants of Metal Ion Selectivity in Proteins. *Chemical Reviews*, **114**, 538-556. <http://dx.doi.org/10.1021/cr4004665>
- [5] Dudev, T., Lin, Y.-L., Dudev, M. and Lim, C. (2003) First-Second Shell Interactions in Metal Binding Sites in Proteins: A PDB Survey and DFT/CDM Calculations. *Journal of the American Chemical Society*, **125**, 3168-3180. <http://dx.doi.org/10.1021/ja0209722>
- [6] Jernigan, R., Raghunathan, G. and Bahar, I. (1994) Characterization of Interactions and Metal-Ion Binding Sites in Proteins. *Current Opinion in Structural Biology*, **4**, 256-263. [http://dx.doi.org/10.1016/S0959-440X\(94\)90317-4](http://dx.doi.org/10.1016/S0959-440X(94)90317-4)
- [7] Marcus, Y. (1988) Ionic Radii in Aqueous Solutions. *Chemical Reviews*, **88**, 1475-1498. <http://dx.doi.org/10.1021/cr00090a003>
- [8] Dudev, M., Wang, J., Dudev, T. and Lim, C. (2006) Factors Governing the Metal Coordination Number in Metal Complexes from Cambridge Structural Database Analyses. *Journal of Physical Chemistry B*, **110**, 1889-1895. <http://dx.doi.org/10.1021/jp054975n>
- [9] Dudev, T. and Lim, C. (2011) Competition between Li⁺ and Mg²⁺ in Metalloproteins. Implications for Lithium Therapy. *Journal of the American Chemical Society*, **133**, 9506-9515. <http://dx.doi.org/10.1021/ja201985s>
- [10] Dudev, T. and Lim, C. (2009) Determinants of K⁺ vs Na⁺ Selectivity in Potassium Channels. *Journal of the American Chemical Society*, **131**, 8092-8101. <http://dx.doi.org/10.1021/ja900168k>