Universitat Autònoma de Barcelona

Understanding the interaction of metal complexes with their biomolecular targets

Giuseppe Sciortino

Universitat Autònoma de Barcelona

UNIVERSITÀ DEGLI STUDI DI SASSARI

Understanding the interaction of metal complexes with their biomolecular targets

An integrated approach

Giuseppe Sciortino

Dissertation submitted for the Degree of *Philosophiae Doctor*

UAB Director
Prof. Jean-Didier Maréchal

UniSS Director
Prof. Eugenio Garribba

-2019-

Understanding the interaction of metal complexes with their biomolecular targets

An integrated approach

Memoria presentada per aspurar al Grau de Doctor per

Giuseppe Sciortino

Vist i plau,

UAB Director
Prof. Jean-Didier Maréchal

UniSS Director
Prof. Eugenio Garribba

The empirical basis of objective science has thus nothing 'absolute' about it. Science does not rest upon solid bedrock. The bold structure of its theories rises, as it were, above a swamp. It is like a building erected on piles. The piles are driven down from above into the swamp, but not down to any natural or 'given' base; and if we stop driving the piles deeper, it is not because we have reached firm ground. We simply stop when we are satisfied that the piles are firm enough to carry the structure, at least for the time being. (Karl Popper, The Logic of Scientific Discovery, 1934)

List of Publications

[1] Sciortino G., Lledos A., and Vidossich P. Bonding Rearrangements in Organometallic Reactions: from Orbitals to Curly Arrows. *Dalton Trans.*, **2019**, 48, 15740–15752.

[2] Sciortino G., Sanna D., Ugone V., Maréchal J.-D., Alemany-Chavarria M., and Garribba E. Effect of Secondary Interactions, Steric Hindrance and Electric Charge on the Interaction of $V^{IV}O$ Species with Proteins. *New J. Chem.*, **2019**. DOI: 10.1039/C9NJ01956A.

[3] Sánchez-Aparicio J.-E., Sciortino G., Herrmannsdoerfer D. V., Chueca P. O., Pedregal J. R.-G., and Maréchal J.-D. Gpathfinder: Identification of Ligand–Binding Pathways by a Multi–Objective Genetic Algorithm. *Int. J. Mol. Sci.*, **2019**, 20(13), 3155.

[4] Singh A. K., Usman M., Sciortino G., Garribba E., and Rath S. P. Through–Space Spin Coupling in a Silver(II) Porphyrin Dimer upon Stepwise Oxidations: $Ag^{II}\cdots Ag^{II}$, $Ag^{II}\cdots Ag^{III}$, and $Ag^{III}\cdots Ag^{III}$ Metallophilic Interactions. *Chem. Eur. J.*, **2019**, 25(43), 10098–10110.

[5] Banerjee A., Dash S. P., Mohanty M., Sanna D., Sciortino G., Ugone V., Garribba E., Reuter H., Kaminsky W., and Dinda R. Chemistry of Mixed–Ligand Oxidovanadium(IV) Complexes of Aroylhydrazones Incorporating Quinoline Derivatives: Study of Solution Behavior, Theoretical Evaluation and Protein/DNA Interaction. *J. Inorg. Biochem.*, **2019**, 199, 110786.

[6] Ugone V., Sanna D., Sciortino G., Maréchal J.-D., and Garribba E. Interaction of Vanadium(IV) Species with Ubiquitin: A Combined Instrumental and Computational Approach. *Inorganic Chemistry*, **2019**, 58(12), 8064–8078.

[7] Sciortino G., Sanna D., Ugone V., Maréchal J.-D., and Garribba E. Integrated ESI–MS/EPR/Computational Characterization of the Binding of Metal Species to Proteins: Vanadium Drug–Myoglobin Application. *Inorg. Chem. Front.*, **2019**, 6, 1561–1578.

[8] Sciortino G., Garribba E., Rodriguez-Guerra Pedregal J., and Marechal J.-D. Simple Coordination Geometry Descriptors Allow to Accurately Predict Metal–Binding Sites in Proteins. *ACS Omega*, **2019**, 4(2), 3726–3731.

[9] Alonso-Cotchico L., Sciortino G., Vidossich P., Rodriguez-Guerra Pedregal J., Drienovska I., Roelfes G., Lledos A., and Marechal J.-D. Integrated Computational Study of the Cu–Catalyzed Hydration of Alkenes in Water Solvent and into the Context of an Artificial Metallohydratase. *ACS Catal.*, **2019**, 9(5), 4616–4626.

[10] Di Meo T., Kariyawasam K., Ghattas W., Valerio-Lepiniec M., Sciortino G., Marechal J.-D., Minard P., Mahy J.-P., Urvoas A., and Ricoux R. Functionalized Artificial Bidomain Proteins Based on an α–Solenoid Protein Repeat Scaffold: A New Class of Artificial Diels–Alderases. *ACS Omega*, **2019**, 4(2), 4437–4447.

[11] Peña Q., Lorenzo J., Sciortino G., Rodríguez-Calado S., Maréchal J.-D., Bayón P., Simaan A. J., Iranzo O., Capdevila M., and Palacios Ò. Studying the Reactivity of "old" Cu(II) Complexes for "novel" Anticancer Purposes. *J. Inorg. Biochem.*, **2019**, 195, 51–60.

[12] Sciortino G., Sánchez-Aparicio J.-E., Pedregal J. R.-G., Garribba E., and Maréchal J.-D. Computational Insight into the Interaction of Oxaliplatin with Insulin. *Metallomics*, **2019**, 11(4), 765–773.

[13] Cabré A., Cabezas-Giménez J., Sciortino G., Ujaque G., Verdaguer X., Lledos A., and Riera A. Mild Iridium–Catalysed Isomerization of Epoxides. Computational Insights and Application to the Synthesis of β-Alkyl Amines. *Adv. Synt. Catal.*, **2019**, 361(15), 3624–3631.

[14] Sciortino G., Lubinu G., Maréchal J.-D., and Garribba E. DFT Protocol for EPR Prediction of Paramagnetic Cu(II) Complexes and Application to Protein Binding Sites. *Magnetochemistry*, **2018**, 4(4), 55.

[15] Salomo E., Gallen A., Sciortino G., Ujaque G., Grabulosa A., Lledós A., Riera A., and Verdaguer X. Direct Asymmetric Hydrogenation of N–Methyl and N–Alkyl Imines with an Ir(III)H Catalyst. *J. Am. Chem. Soc.*, **2018**, 140(49), 16967–16970.

[16] Cabre A., Sciortino G., Ujaque G., Verdaguer X., Lledos A., and Riera A. Iridium–Catalyzed Isomerization of N–sulfonyl Aziridines to Allyl Amines. *Org. Lett.*, **2018**, 20(18), 5747–5751.

[17] Sciortino G., Lihi N., Czine T., Maréchal J.-D., Lledós A., and Garribba E. Accurate Prediction of Vertical Electronic Transitions of Ni(II) Coordination Compounds via Time Dependent Density Functional Theory. *Int. J. Quantum Chem.*, **2018**, 118(16), e25655.

[18] Lopez S., Muñoz, Couce-Rios A., Sciortino G., Lledos A., and Ujaque G. Mechanistic Insights on the Hydration of Terminal and Internal Allenes Catalyzed by $[(NHC)Au]^+$. *Organometallics*, **2018**, 37(20), 3543–3551.

[19] Pedregal J. R.-G., Funes-Ardoiz I., Sciortino G., Sánchez-Aparicio J.-E., Ujaque G., Lledós A., Maréchal J.-D., and Maseras F. GARLEEK: Adding an Extra Flavor to ONIOM. *J. Comput. Chem.*, **2019**, 40(2), 381–386.

[20] Sanna D., Ugone V., Sciortino G., Parker B. F., Zhang Z., Leggett C. J., Arnold J., Rao L., and Garribba E. $V^{IV}O$ and V^{IV} Species Formed in Aqueous Solution by the Tridentate Glutaroimide–Dioxime Ligand – An Instrumental and Computational Characterization. *Eur. J. Inorg. Chem.*, **2018**, 2018(17), 1805–1816.

[21] Rodríguez-Guerra J., Alonso-Cotchico L., Sciortino G., Lledós A., and Maréchal J.-D. Computational Studies of Artificial Metalloenzymes: From Methods and Models to Design and Optimization. *Artificial Metalloenzymes and MetalloDNAzymes in Catalysis: From Design to Applications*, **2018**.

[22] Gómez-González J., Peña D. G., Barka G., Sciortino G., Maréchal J.-D., Vázquez López M., and Vázquez M. E. Directed Self-Assembly of Trimeric DNA-Binding Chiral Miniprotein Helicates. *Front. Chem.*, **2018**, 6, 520.

[23] Sciortino G., Sanna D., Ugone V., Lledos A., Marechal J.-D., and Garribba E. Decoding Surface Interaction of $V^{IV}O$ Metallodrug Candidates with Lysozyme. *Inorg. Chem.*, **2018**, 57(8), 4456–4469.

[24] Sciortino G., Garribba E., and Marechal J.-D. Validation and Applications of Protein–Ligand Docking Approaches Improved for Metalloligands with Multiple Vacant Sites. *Inorg. Chem.*, **2018**, 58(1), 294–306.

[25] Sciortino G., Rodríguez-Guerra Pedregal J., Lledós A., Garribba E., and Maréchal J.-D. Prediction of the Interaction of Metallic Moieties with Proteins: An Update for Protein–Ligand Docking Techniques. *J. Comput. Chem.*, **2018**, 39(1), 42–51.

[26] Sanna D., Ugone V., Sciortino G., Buglyó P., Bihari Z., Parajdi-Losonczi P. L., and Garribba E. $V^{IV}O$ Complexes with Antibacterial Quinolone Ligands and their Interaction with Serum Proteins. *Dalton Trans.*, **2018**, 47(7), 2164–2182.

[27] Sciortino G., Sanna D., Ugone V., Micera G., Lledos A., Marechal J.-D., and Garribba E. Elucidation of Binding Site and Chiral Specificity of Oxidovanadium Drugs with Lysozyme through Theoretical Calculations. *Inorg. Chem.*, **2017**, 56(21), 12938–12951.

[28] Rodríguez-Guerra Pedregal J., Sciortino G., Guasp J., Municoy M., and Maréchal J.-D. Gaudimm: A Modular Multi–Objective Platform for Molecular Modeling. *J. Comput. Chem.*, **2017**, 38(24), 2118–2126.

[29] Lihi N., Godó A. J., Sciortino G., Garribba E., and Várnagy K. Tridentate (O, N, O) Ligands as Potential Chelator Compounds for Iron Overload. *Polyhedron*, **2017**, 123, 192–205.

[30] Sanna D., Sciortino G., Ugone V., Micera G., and Garribba E. Nonoxido V^{IV} Complexes: Prediction of the EPR Spectrum and Electronic Structure of Simple Coordination Compounds and Amavadin. *Inorg. Chem.*, **2016**, 55(15), 7373–7387.

Abstract

Decoding the interaction of transition metal complexes with proteins is fundamental in biology, pharmacy, medicinal chemistry fields, as well as in the design of artificial metalloenzymes. X-ray crystallography and Nuclear Magnetic Resonance (NMR) methods provide with explicit structural description of the system although their application is far from trivial. Other spectroscopic methods give insight on the protein region where the metal species is bound or on the amino acid type involved in the metal coordination. Mass spectrometry based techniques can determine the metal moieties-protein stoichiometry, and often suggest the coordinating amino acids. These instrumental techniques though are not able to provide with a complete three dimensional description of the system.

This Ph.D. was aimed at expanding molecular modeling applicability to the prediction of the interaction of metallo-ligands with proteins. The thesis has been built on two major pillars. On the one hand protein-ligand docking techniques have been updated to deal with protein-metal coordination bonds formation and on the other, the resulting tools have been applied to a series of practical cases. These have been focused on the study of the interaction of vanadium based insulin-enhancing, and platinum based anti-tumor metallodrugs with their biological target, and the characterization of a copper based artificial metalloenzyme with Diels-Alderase activity. Moreover, a new integrated approach based on the combination of molecular modeling with the EPR and ESI-MS instrumental techniques, is presented. This part of the work provides with major proof-of-concept on the power of coupled experimental and theoretical approaches for the rational design of new metallodrugs as well as act as a guide in a large number of fields in bioinorganic chemistry.

Riassunto

Cartterizzare l'interazione tra i complessi dei metalli di trasizione e le proteine è fondamentale in diversi campi della ricerca scientifica come la biologia, la farmacologia, la chimica farmaceutica e il disegno di nuovi metalloenzimi artificiali. La diffrattometria a raggi-X e la Risonanza Magnetica Nucleare (NMR) possono, potenzialmente, fornire un'accurata descrizione strutturale del sistema, nonostante possano presentare svariate difficoltà che spesso ne precludono l'applicazione. Altre tecniche spettroscopiche sono in grado di fornire importanti informazioni circa la regione della proteina in cui si trova legato il complesso metallico o, ancora, sul tipo di donatori coinvolti nel legami di coordinazione metallo-proteina. Tecniche basate sulla spettrometria di massa sono in grado di determinare il rapporto stechiometrico metallo/proteina degli addotti formati però raramente possono dare indicazioni sull'amminoacido specifico che coordina il metallo.

L'obiettivo principale di questo lavoro di tesi è stato affrontare questa problematica da un punto di vista computazionale. A tal fine si sono estese le principali tecniche di modellistica molecolare, usate fino ad ora principalmente su sistemi organici, allo studio e alla predizione dell'interazione tra complessi di metalli di transizione con le proteine. La tesi presenta due diversi livelli, da una parte lo sviluppo metodologico di nuove tecniche di *docking* capaci di descrivere il legame di coordinazione proteina-metallo, e dall'altra la applicazione di queste tecniche ad una serie di casi di interesse pratico. Nello specifico lo studio dell'interazione di metallofarmaci insulino-mimetici a base di vanadio e antitumrali a base di platino con i loro *target* biologici e la caratterizzazione di un enzima artificiale, con attività *Diels-Alderasi* basata su un cofattore di rame. Viene proposto, inoltre, un protocollo che integra la modellisca molecolare con tenciche di spettroscopia EPR e spettrometria ESI-MS. Questo nuovo approccio mostra elevate potenzialità di applicazione nella progettazione razionale di nuovi metallofarmaci e metallo-enzimi oltre che nella chimica bioinorganica in generale.

Resumen

Caracterizar la interacción entre complejos de metales de transición y proteínas es fundamental en campos de investigación como la farmacología y la química médica y para el diseño de metaloenzimas artificiales. La difracción de rayos-X y la resonancia magnética nuclear (RMN) proporcionan una descripción estructural completa del sistema, aunque pueden presentarse dificultades que impidan su aplicación. Otras técnicas espectroscópicas facilitan información de la región en la que el complejo metálico se encuentra enlazado a la proteína, o del tipo de donadores que dan lugar al enlace, pero no permiten una resolución estructural explicita del sistema. Técnicas basadas en la espectrometría de masas son capaces de determinar la relación estequiométrica metal/proteína, pero raramente pueden dar indicaciones acerca del aminoácido específico que coordina el metal.

El objetivo general de la presente tesis es atacar esta problemática mediante métodos computacionales de simulación. Para ello, las principales técnicas de modelización molecular, aplicadas hasta ahora principalmente a sistemas orgánicos, se han extendido al estudio y la predicción de la interacción entre complejos de metales de transición y proteínas. La tesis tiene una doble vertiente, por un lado, el desarrollo metodológico de nuevas técnicas de *docking* capaces de representar el enlace de coordinación proteína-metal, y por el otro la aplicación de estas técnicas a una serie de casos de interés práctico. Como aplicaciones se ha estudiado la interacción de metalofármacos insulino-miméticos a base de vanadio y antitumorales a base de platino con sus *targets* biológicos y la caracterización de una enzima artificial con actividad *Diels-Alderasa*, basada en un cofactor de cobre. Se propone además un protocolo integrado para el estudio de interacciones proteína-metal combinando la modelización molecular con técnicas instrumentales de espectroscopia EPR y espectrometría ESI-MS. Esta nueva aproximación muestra elevadas potencialidades de aplicación en el diseño racional de nuevos metalofármacos y metaloenzimas, y en el campo general de la química bioinorgánica.

Contents

Chapter 1

Introduction

1.1 Metals, Chemistry and Biology

With more than 30% of the genome encoding for metal containing biomolecules, the study of the interaction between metal ions and proteins is of fundamental aspect in biology, pharmacy and medicine. [1–3] Metal ions show a wide range of biological functions, from structural (Ca, Zn, Si), signalling (Ca), electron transfer (Fe, Cu, Mo), to acid base (Zn, Fe, Ni, Mn) or redox catalysis (Mn, V, Fe, Co, Ni, Cu, W). Metals can be strongly bound to the receptor forming metalloproteins (*e.g.* Fe, Cu or Zn in cytochrome, blue proteins and Zn fingers), or weakly, playing an activator role on metal dependent enzyme (*e.g.* Ca, Mg, and Mn in kinases or adenylate cyclase). The kind of interaction governing the metal to protein binding process depends on the metal nature, and moves to pure electrostatic for main group metals, to proper coordination bond in transition metals (TM). The most common protein donors are represented by amino acid side chains and in particular, depending on the Hard or Soft character of the metal can be nitrogen atoms of histidine, oxygen atoms of glutamate, aspartate, tyrosinate or asparagine and glutamine residues, sulphur atoms of methionine and cysteine. Moreover, amide groups of the backbone through O- or deprotonated N-donors are sometimes implied in coordination. [4] Additionally, metal ions are also involved in diseases such as neurodegenerative pathologies with amyloid plaques formation, transmissible spongiform encephalopathies, the Lewy bodies of Parkinson's disease and many others. [5–11]

Over the last years, the chemical research was inspired by the metals biological functions and applied in many fields like therapy (*e.g.* metallodrugs), diagnosis (*e.g.* biosensors) and industrial processes (artificial metalloenzymes, ArM). [2, 3, 12–22]

In this field the role of Computational Chemistry has reached an increasing role to support and complement experimental results. [23–27]

1.1.1 Metals in Living Organisms

About 24 periodic table's elements are currently recognized as essential although the precise requirements spread between μg/kg to mg/kg within different organisms. The organic elements H, C, N, and O, followed by the alkaline and alkaline earth metals Na, K, Mg, Ca and the main group elements Cl, F, P, S, Si and Se. Furthermore, *trace elements* such as the main group Br, Si and Sn or the transition metals Mn, Fe, Cu, Zn, Co, Cr, Mo V, Ni are also defined as essential for several organisms. [28–31] The definition *essential* is given for those elements for which a physiological deficiency appears if removed, and disappears if reintroduced in the diet and/or their specific biological function is well characterized. Among the transition metals mentioned above, Mg, Mn, Fe, Co, Cu, Zn, Mo, V, Ni are today considered essential or beneficial for humans despite for the two latter, a precise biological function still elusive. [32, 33]

As reported in Table 1.1 for several selected elements, the amount of metal ions at service of life can be related to their abundance in earth and oceans that can vary over several orders of magnitude. [4, 34]

Element	**Crust(ppm)**	**Sea Water(μM)**	**Blood Plasma(μM)**	**Cell/Tissue(μM)**
Ca	4×10^4	1×10^4	2×10^3	1×10^3
Cd	0.2	1×10^{-3}		
Co	25	2×10^{-5}	$2.5x10^{-5}$	
Cu	55	4×10^{-3}	$8 - 24$	~ 68
Fe	5×10^4	1×10^{-3}	22	$0.001 - 10$
K	3×10^4	1×10^4	4×10^3	1.5×10^5
Mg	2×10^4	5×10^4	500	$9x10^3$
Mn	950	5×10^{-4}	0.1	180
Mo	1.5	0.1		5×10^{-3}
Na	3×10^4	5×10^5	1×10^5	1×10^4
Ni	75	8×10^{-3}	0.04	2
V	135	0.03	0.07	$0.5 - 30$
W	1.5	5×10^{-3}		
Zn	70	0.01	17	180

Table 1.1: Average relative abundance of selected elements in earth's crust, sea water, mammalian blood plasma and mammalian cells or tissues from ref. [34]

For mammalian cells the source of metal ions is the blood plasma, in which the levels of metal ions are controlled within a narrow range of concentrations by regulating absorption and/or excretion of metal ions from the body. Cells that obtain their metal ions from plasma accumulate most of these elements to much higher levels than the plasma itself. The observation that cells must often acquire metal ions against a concentration gradient demonstrates the importance of efficient metal ion uptake systems.

The chemical properties of metal ions can greatly affect their availability to organisms, among them redox chemistry, hydrolysis, solubility, chelation of the free ion and ligand exchange rates must be mentioned as will be discussed in the following section for case of vanadium.

It must be highlighted that, as reported in Table 1.2, *Soft* metals rarely are essential for life, moreover they are often toxic (Hg^{+} , Hg^{2+} , Cd^{2+} , Tl^{+}).

Hard	**Borderline**	**Soft**
	Acids	
H^{+}, Li^{+}, Na^{+}, K^{+}, Be^{2+}, Mg^{2+}, Ca^{2+}, Sr^{2+}, Sc^{3+}, Al^{3+}, Ga^{3+}, In^{3+}, Cr^{3+}, Co^{3+}, Fe^{3+}, Ir^{3+}, La^{3+}, Si^{4+}, Ti^{4+}, Zr^{4+}, Th^{4+}, U^{4+}, VO^{2+}, UO_2^{2+}, $BeMe_2$, BF_3, BCl_3, $B(OR)_3$, $AlCl_3$, $AlMe_3$	Fe^{2+}, Co^{2+}, Ni^{2+}, Cu^{2+}, Zn^{2+}, Sn^{2+}, Pb^{2+}, Rh^{3+}, $B(CH_3)_3$, R_3C^{+}, SO_2 , NO^{+}	Cu^{+}, Ag^{+}, Au^{+}, Hg^{+}, Cs^{+}, Tl^{+}, Hg^{2+}, Pd^{2+}, Cd^{2+}, Pt^{2+}, MoO_2^{2+}
	Bases	
H_2O, OH^{-}, F^{-}, CH_3CO^{2-}, PO_4^{3-}, SO_4^{2-}, CO_3^{2-}, NO^{3-}, ClO^{4-}, O^{2-}, ROH, RO^{-}, R_2O, NH_3, RNH_2, N_2H_4	Aniline, pyridine, N_3^{-}, Cl^{-}, Br^{-}, NO_2^{-}, SO_3^{2-}, N_2	S^{2-}, RSH, RS^{-}, R_2S, I^{-}, CN^{-}, SCN^{-}, $S_2O_3^{-}$, R_3P, R_3 As, $(RO)_3P$, RNC, CO, C_2H_4 , C_6H_6, R^{-}, H^{-}

Table 1.2: Hard and Soft acids and bases from ref. [35]

Considering the HSAB (Hard and Soft Acids and Bases) [36, 37] principles, these *Soft* ions exhibit high affinity for *Soft* bases, thus, also in a very low concentration, are able to disrupt the structure and function of enzymes through hydrogen bonds or disulphure bond breaking by forming S-metal bonds. None of the typical *Soft* metal ion is essential, except Cu^{+} which can be present during redox reactions involving Cu^{2+} ions or intra-cellularly tightly bound by especially designed chaperons and metalloenzymes.

On the contrary, many of the *borderline* transition metal ions are essential and their appreciable Lewis acidity make them able to selective coordinate some fundamental ligands (*e.g.* O_2) or to improve the acidity or reactivity of ligands. Moreover, some of them show several accessible oxidation states, sometimes differing only by one electron, allowing to catalyze redox enzymatic reactions. [38]

Metal ions or their compounds have several biological functions which can be summarized as follows: [30, 35, 38]

i) participation in degradation, formation or metabolism of organic compounds in several enzymes (Zn^{2+})

ii) transfer of electrons in redox pairs, also by stabilizing "unusual" oxidation states of redox-active metals: Fe(II), Fe(III), Fe(IV); Cu(I), Cu(II); Mn(II), Mn(III), Mn(IV); Mo(IV), Mo(V), Mo(VI); Co(I), Co(II), Co(III)

iii) catalysis of a wide class of reactions under physiological conditions: most of the transition metals are able to provide unpaired electrons and to concurrently donate electronic charge through a π back donation bonding.

Some general characteristics as electronic d configuration, coordination number (CN) and geometry, donor preferences and biological functions of transition metals are reported in Table 1.3.

Metal Ion	C.N.	Geometry	Donors	Functions
$Mn^{2+}(d^5)$	6	Octahedral	O-Carboxylate, Phosphate, N-Imidazole	Structure in oxidases, photosynthesis
$Mn^{3+}(d^4)$	6	Tetragonal	O-Carboxylate, Phosphate, Hydoxime	Structure in oxidases, photosynthesis
	6	Octahedral	O-Carboxylate, Phenolate, Alkoxide, Oxide, N-imidazole, Porphyrins	
$Fe^{2+}(d^6)$	4	Tetrahedral	S-Thiolate	Electron transfer, nitrogen fixation in nitrogenases
	6	Octahedral	O-Carboxylate, Phenolate, Alkoxide, S-Thiolate, Oxide, N-Imidazole, Porphyrins	Electron transfer in oxidases, O_2 transport
$Fe^{3+}(d^5)$	4	Tetrahedral	S-Thiolate	Electron transfer, nitrogen fixation in nitrogenases
	6	Octahedral	O-Carboxylate, Phenolate, Alkoxide, Oxide, N-Imidazole, Porphyrins	Electron transfer in oxidases
$Co^{+}(d^8)$	6	Octahedral	O-Carboxylate, N-Imidazole	Alkyl group transfer in vitamin B_{12s}
$Co^{2+}(d^7)$	4	Tetrahedral	S-Thiolate, N-Imidazole	Alkyl group transfer, oxidases
	6	Octahedral	O-Carboxylate, N-Imidazole	Alkyl group transfer in vitamin B_{12r}
$Co^{3+}(d^6)$	6	Octahedral	S-Thiolate, Thioether, N-Imidazole	Alkyl group transfer in vitamin B_{12}
$Cu^{+}(d^{10})$	4	Tetrahedral	S-Thiolate, N-Imidazole	Electron transfer in Type I blue copper proteins and Type III heme-copper oxidases
$Cu^{2+}(d^9)$	4	Tetrahedral	S-Thiolate, N-Imidazole	Electron transfer in Type I blue copper proteins and Type III heme-copper oxidases
	4	Square Planar	O-Carboxylate, N-Imidazole	Type II copper in oxidases
	5	Square Pyramidal	O-Carboxylate, N-Imidazole	Type II copper oxidases, hydrolases
	6	Tetragonal	O-Carboxylate, N-Imidazole	Type III copper hydroxylases, dioxygen transport in hemocyanin
$Zn^{2+}(d^10)$	4	Tetrahedral	O-Carboxylate, Carbonyl, S-Thiolate, N-Imidazole	Structure in zinc fingers, gene regulation, anhydrases, dehydrogenases
	5	Square Pyramidal	N-Imidazole, O-Carboxylate, Carbonyl,	Structure in hydrolases, peptidases

Table 1.3: Ligands and geometries preferred by different metal ions. From ref. [35]

1.1.2 Metals in Medicine

Medicinal chemistry has arisen as a field covering the design and synthesis of novel drugs based on knowledge related to their function at the molecular level. This new prospect enabled the development of metal coordination compounds acting as useful metallodrugs which interact with a specific molecular target (pharmacodynamics), and being capable of reaching that target (pharmacokinetics). As the main difference respect to many organic drugs, which follow conventional design rules, metallodrugs are in general pro-drugs that undergoes *in vivo* transformations before reaching their target sites. As mentioned in the previous paragraph the bio-transformations can include redox reactions of the metal ion, ligand substitution, or reactions of organic functionalities of the ligands far from the metal. The exact knowledge of the speciation and mechanisms of action of these compounds is a fundamental aspect in metallodrugs design and it requires drawing together the reactivity of the metal and their ligands.

Metals were proposed in various fields of medicine and several health problems as bacterial, viral (particularly HIV), and parasitic infections such as malaria, are being investigated on metal-based medicine research. [3, 39–44] Promising applications were pointed out in the treatment of several diseases like cardiovascular, age-related inflammatory diseases, neurological diseases (*e.g.*, Parkinson's, Alzheimer's), cancer, diabetes, and arthritis. [45–50] As an example of this family of compounds can be mentioned the antiulcer agent Bi^{III}−citrate (De NolTM), the antiarthritic triethylphosphine Au^{I} complex (auranofin), the anticancer derivatives of cisplatin such as carboplatin and oxaliplatin, the antitumor Ru^{III} NAMI-A and the Gd^{III} complexes used for contrast enhancement in magnetic resonance imaging (MRI). [2, 3, 12–14, 17]

In 1985, an orally active gold drug named auranofin ((2,3,4,6-tetra-O-acetyl-1-thio-κS-β-D-glucopyranosato-S)(triethylphosphine)gold(I), Ridaura) was introduced for the treatment of rheumatoid arthritis. [17,51,52] Auranofin was shown to be also highly cytotoxic to tumor cells in culture and is currently being evaluated for the treatment of chronic lymphocytic leukemia (CLL) in phase II clinical trials. [53] Other Au^{I} species based on the moiety $[Au(PEt_3)]^{+}$ for example, $[Au(Et_3P)Cl$, $[Au(Et_3P)CH_3]$, $[Au(Et_3P)CN]$, $[Au(Et_3P)SCH_3]$, $Au(Et_3P)SCN]$, $[Au(Et_3P)SPh]$ were evaluated for *in vitro* cytotoxicity against both B16 melanoma cells and P388 leukemia cells and *in vivo* anti-tumor activity against P388 leukemia in mice. [54]

Cisplatin (Figure 1.1 a) has a central role in cancer chemotherapy, especially for testicular cancer for which the overall cure rate exceeds 90%, and is nearly 100% for early-stage disease. [55, 56] Cisplatin have been in widespread use for many years to treat several other forms of cancer, including ovarian, cervical, head and neck, and non-small-cell lung cancer, usually combined with radiation or other medications. Cisplatin is believed to induce apoptosis in cancer cells by

covalently modifying the DNA. [57] The treatment is limited, however, by side effects including nephrotoxicity, emetogenesis and neurotoxicity. [58]

The search for anticancer platinum compounds to replace cisplatin – which has several side effect such as kidney toxicity, nausea, vomiting, tinnitus and bone marrow suppression - has brought to the second generation Pt drugs with lower toxicity. Among others, these compounds include carboplatin and oxaliplatin (Figure 1.1 b and c). [16, 17, 59] Oxaliplatin (whose clinically used infusion is named Eloxatin), [Pt(dach)(ox)], where dach is 1*R*,2*R*-diaminocyclohexane and ox is oxalate, is an anticancer drug which has been approved for clinical use in the European Union in 1999 and in the United States in 2002. [60, 61] Oxaliplatin is especially effective against metastatic colorectal cancer and is much less nephro- and ototoxic than cisplatin. In addition, oxaliplatin can be used to treat cisplatin-resistant tumors. Oxaliplatin contains the labile dianionic oxalate ligand and in the blood gives $[Pt(dach)]^{+}$, with two free coordination positions, that can be attacked by serum bioligands, among which proteins.

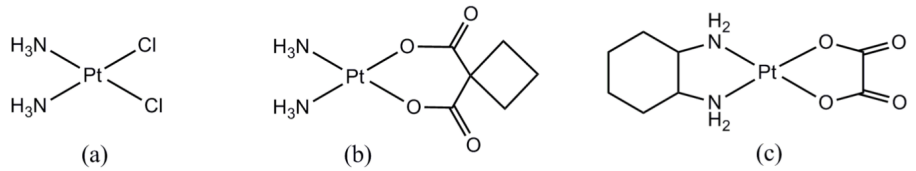

Figure 1.1: Structure of cisplatin (a), carboplatin (b), and oxaliplatin (c)

Large number of ruthenium compounds have been synthesized and tested for their anticancer properties. Two of the them, named NAMI-A and KP1019, which contain Ru^{III}, are in clinical trials. [62, 63] The experimental evidence that these complexes could be reduced by bioligands of the orgainism, suggested that the reduction of Ru^{III} pro-drug to Ru^{II} is a fundamental step to achieve their pharmacological activity. [64, 65] Therefore, the anticancer potential of Ru^{II} species has been examined and, among these compounds, generically indicated generically as RAPTA species, $[Ru^{II}Cl_2(\eta-6\text{-arene})(pta)]$, where pta is 1,3,5-triaza-7-phosphaadamantane, have shown very promising action and possible application. [66]

1.1.3 Metallo Enzymes

All reactions that take place in biological environment have to be carried out under strict control, they must only happen as they are required and where they are required. The way for which organisms obtain this total control is employing reactions that can only proceed under enzymatic catalysis that increases the

reaction rate up to a factor of 10^{17}. [67] More than half of all enzymes have metal ions in their structure. Metalloenzymes range in size from about 80 amino acids (as in the case of cytochrome c_{551}) up to hundred of kDa (such as for glutamine synthetase). In most cases, the metals are essential to the catalytic action of the enzyme and are often at the active site where the substrate for the biochemical reaction is bound. Several enzymes catalyzed reaction require a given co-catalyst such as in the case of the coenzyme B_{12} for which a Co-containing cofactor is required. In general, the metal ions can: i) act as an anchoring for the substrates through coordination bond or electrostatic interactions; ii) enhance the acidity of the coordinated substrate; iii) act as a redox center: iv) play a simple structural role.

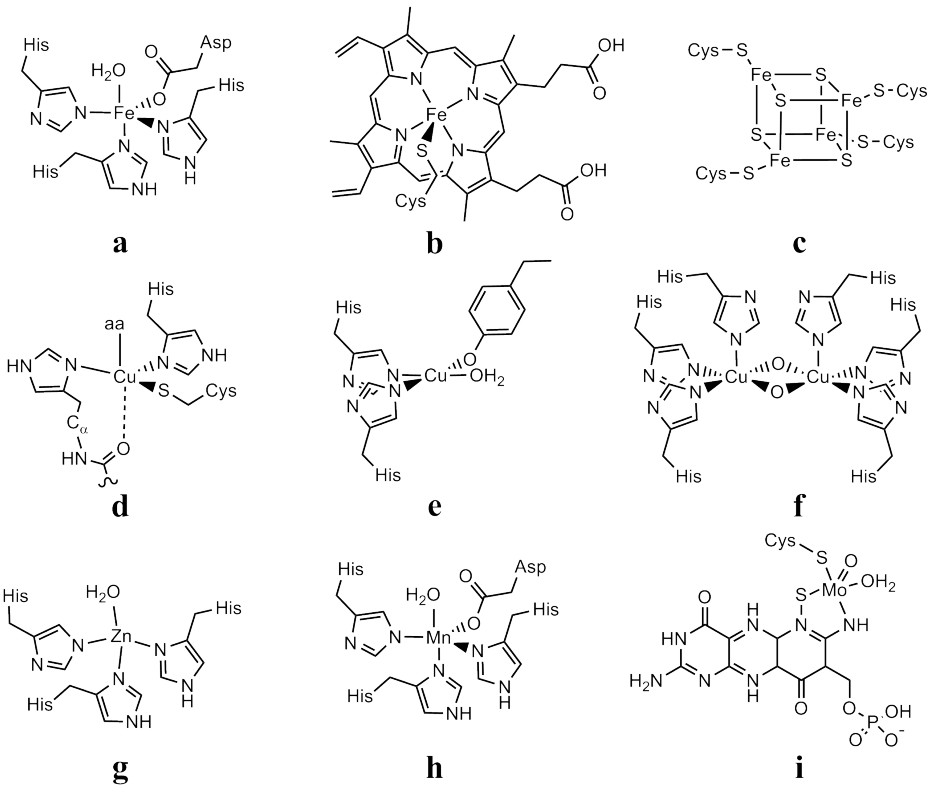

Figure 1.2: Scheme of selected metalloenzymes active site: a) Fe-superoxide dismutase, b) heme P450, c) Fe-S simplest cluster, d) azurin, e) galactose oxidase, f) hemocyanin, g) Zn alcohol dehydrogenases, h) Mn superoxide dismutase and i) molybdopterin

Mononuclear Fe proteins are in general oxidoreductases such as Fe-superoxide dismutase (Figure 1.2, a), lipoxygenases, nitrile hydratase *etc.* or electron transfer enzymes such as rubredoxins or photosynthetic centers. The most common Fe-coordinating donors are N-Imidazolic of histidines, S-Thiolate of cystein or N-amide of the backbone. The active site of dinuclear oxidoreductaxe such as alkenehydroxylase, methane monooxygenase, and phenolhydroxylase is formed by a coupled binuclear iron center connected by several bridging O-carboxylate and N-Imidazolic histidine residues.

The prostetic group of hemoproteins is represented by an iron porphyrin in which the axial coordination positions are used for the scaffold and substrate coordination (Figure 1.2, b). The most common hemoproteins are oxidoreductases, catalases or peroxidases, and electron transfer proteins such as cytochromes. [68,69] Iron-sulfur proteins are characterized by iron-sulfur (Fe-S) clusters containing sulfide-bridged di-, tri-, and tetra-iron centers with variable oxidation states (Figure 1.2, c), the most common Fe-S enzymes are ferredoxins, NADH hydrogenases, dehydrogenases, cytochrome c reductases and nitrogenases. Fe-S proteins play a central role in respiratory chains, of either bacteria and archaea or in the mitochonria of eucaryot organisms. Additional functions are stereo-specific isomerization of citrate to isocitrate (aconitase) or the redox-dependent gene regulation. [67]

Copper containing enzymes are involved in activation processes and electron transport. *Type I* centers such as ascorbate oxidase, ceruloplasmin, laccase, nitritereductase, auracyanin, and azurin are characterized by a copper ion coordinated by two N-Imidazolic nitrogens of histidine and a S-Thiolate of cysteine in a trigonal planar arrangement and a variable axial amino acid donor (Figure 1.2, d). *Type II* centers are in a square planar arrangement, coordinated by N or N/O ligands such as in in galactose oxidase (Figure 1.2, e). *Type III* are dinuclear centers, each coordinated by three N-Imidazolic of histidine such as in catechol oxidase, hemocyanins, tyrosinase (Figure 1.2, f). The dinuclear *copper A* center of cytochrome c-oxidase is coordinated by two N-Imidazolic of histidine, S-methionine, O-carbonyl and two S-Thiolate of cysteine bridging the copper atoms. In cytochrome oxidase, *copper B* centers, are coordinated by three N-Imidazolic of histidine. The *four copper center* of nitrous oxide oxidase is coordinated by seven N-Imidazolic of histidine residues. The four ions are bridged by one S-Thiolate. Besides cytochrome oxidase, monoamine-oxidase, laccase, caeruloplasmin, superoxide-dismutase and hemocyanine are other examples of oxygenases or dioxygenases. [70]

Zinc proteins with catalytic activity are generally characterized by the coordination of three side chains donors and a water molecule in a tetrahedral arrangement. The filled *d* orbital of Zn^{2+} provide a redox inertia to the metal, however their rather Lewis acid character is favorable for reactions like proteolysis or hydration of carbon dioxide. Important zinc-containing enzymes are the

alcohol dehydrogenases, metalloproteases and carbonic anhydrase (Figure 1.2, g). Moreover co-catalytic zinc sites are found in superoxide-dismutases, phosphatases, aminopeptidases, and β-lactamases. [71, 72]

Nickel in metalloenzymes occurs as Ni-4Fe-5S as in nickel-containing CO dehydrogenases, coordinated in a tetrapyrrole complex as in in methyl-coenzyme M reductase with glutamine as axial ligand, in dinuclear NiNi or NiFe complexes such as hydrogenases and urease or a $[4\,\mathrm{Fe}{-}4\,\mathrm{S}]^{2+}$ cluster bridged by a S-Thiolate of cysteine to a dinuclear NiNi center as in acetyl-CoA synthase. [73]

Manganese (normally Mn(II)) in metalloenzymes may act as a simple Lewis acid catalyst or can also be oxidized to the +III or +IV which allows redox-cycling reactions. Manganese metal centers may be mono-, bi-, tri-, or tetranuclear. Mn-superoxide dismutases (SODs) exhibit mononuclear centers and are antioxidant metalloenzymes catalyzing the redox disproportionation of the superoxide radical, O_2^- (Figure 1.2, h). The active site manganese is five-coordinate, with the metal ligands arranged in distorted trigonal bipyramidal geometry. Mo and W enzymes catalyze basic metabolic reactions in the nitrogen, sulfur, and carbon cycles. [74–76] With the exception of the nitrogenase cofactor, molybdenum is associated with the heterocyclic pterin derivative (molybdopterin) that contains a mononuclear center, coordinated to the thiols of the cofactor (Figure 1.2, i). Tungsten exhibits a similar coordination to the pterin cofactors. Due to the higher sensitivity of W toward high redox potential, tungsten enzymes are found in (thermophilic) anaerobic bacterial and archaeal microorganisms. [77, 78]

Cobalt is the central metal ion in the tertrapyrrolcorrin ring in the coenzyme B_{12}. The B_{12}-based enzymes catalyze either the transfer of methyl groups between two molecules or isomerization reactions, *e.g.* rearrangements of hydrogen atoms with concomitant exchange of a substituent (*e.g.*, an hydroxylgroup) between two adjacent carbon atoms in a molecule. In non-corrin cobalt enzymes, cobalt is complexed by aminoacids such as in nitrile hydratase. [79]

Artificial Metalloenzymes (ArMs)

Homogeneous transition metal catalysts consist of a metal ion covalently bound to some number of ligands that constitute the first coordination sphere of the metal. Its reactivity is modulated by the nature of the metallic center (redox properties, Lewis acidity), the number and donor/acceptor character of the ligands, and their geometry around the metallic center. In some cases the ligands can actively participate in the chemical transformation (cooperative metal–ligand catalysis, redox non-innocent ligands). The first coordination sphere controls the catalytic activity and selectivity of the transition metal catalyst. Fine-tuning of the ligand substituents brings in selectivity. In this approach it is possible, using very sophisticated ligands, to induce enantioselectivity by forcing an incoming

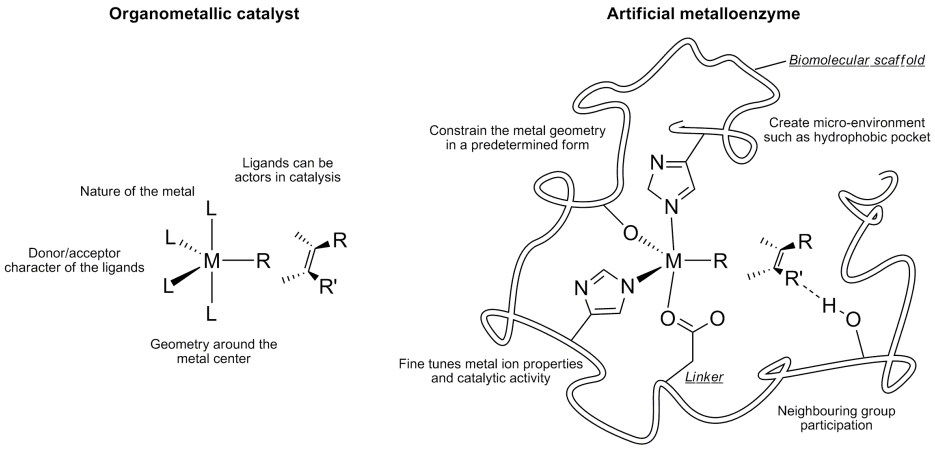

Figure 1.3: Main features of a transition metal catalyst (a) and an artificial metalloenzyme (b). Adapted from ref. [80]

reagent to approach selectively from one prochiral side of the substrate (Figure 1.3).

ArMs are devised to harness the potential of transition metal and enzyme catalysis by combining the best of both worlds in a single entity. The biomolecular scaffold, which in a simplified way may be described as a very big multidentate and modular ligand that accommodates metal binding, [81] introduces a secondary coordination environment that can be readily and extensively fine-tuned. In such entities the first coordination sphere (metal+ligands) provides the catalytic activity, while the second coordination sphere afforded by the biomolecular scaffold confers specificity and selectivity (Figure 1.3). By handling the catalytic center and introducing the proper interactions in the secondary sphere, it is possible to control stereo- and regio-selectivities. Compared with transition metal catalysts, the potential for optimization of ArMs is notably increased. The cofactor and the biomolecule can be optimized independently through synthetic and molecular biology approaches. By chemical procedures, the ligands, their substituents, and the metal itself can be modified. On the other hand, mutating the amino acid sequence could diversify the host protein. This combined optimization strategy, called the chemogenetic approach, yields a diverse library of compounds to be screened for activity and selectivity, and it has proven to be very successful. Despite some pioneering work in the seventies, [82] the introduction of ArMs to catalyze chemical transformations relevant for organic synthesis is very recent. [19, 21, 22, 81, 83, 84]

For instance, regarding asymmetric catalysis, inversion of enantioselectivity

has been achieved by introducing small changes in the bioconjugated systems. In an artificial enantioselective Suzukiase based on the biotin–streptavidin technology, varying the length of the spacer between the biotin anchor and the $P(t\text{-}Bu)_2$ from ethyl to propyl affords opposite enantiomers of a biaryl. [85] The same effect has been obtained by site-directed mutagenesis in the streptavidin (Sav) scaffold of an artificial imine reductase. Upon substitution of Ser112 with either an alanine or a lysine, both enantiomers of salsolidine can be produced. [86]

1.2 Vanadium Biochemistry

Vanadium (element no. 23) with a 0.0001% cosmic abundance is quantitatively comparable to copper and zinc. It is distributed ubiquitously, in soil, fossils (in particular in crude oil), water, air and living organisms. The earth's crust, including the aqua- and atmosphere, contains *ca.* 130 ppm (by mass) of vanadium, making it the 21nd most abundant element in the outermost sphere. [87] The average content in shale, which are particularly rich in vanadium, is 0.012% w/w. More than 120 vanadium-based minerals are known containing the element in cationic and anionic form, and in the oxidation states +III, +IV and +V. In sea water, the average concentration of vanadium is around 30 nM. Vanadium is thus the second most abundant transition element in marine environments, outmatched only by molybdenum (*ca.* 100 nM). [88]

Vandium is an essential element for many biological systems and a trace element in humans. In general can assume oxidation states ranging from -I to +V and, +III, +IV and +V have been found in biological systems.

V^{III} has a $3d^2$ electronic configuration and is normally hexa-coordinated with octahedral geometry. However, its relatively large ionic radius allows it to form hepta-coordinated complexes of distorted pentagonal bipyramidal structure. At biologically relevant concentrations ($< 1\ \mu M$), V^{III} exists as the aqua ion $[V^{III}(H_2O)_6]^{3+}$ only in acidic solutions. At micromolar concentrations, $[V^{III}(H_2O)_5(OH)]^{2+}$ and $[V^{III}(H_2O)_4(OH)_2]^{+}$ are formed with pK_a values of 2.7 and 4.0. [88]

V^{IV} has a $3d^1$ electronic configuration making it an ideal candidate for electron paramagnetic resonance spectroscopy (EPR). In acidic aqueous solution, V^{IV} exists as the oxidovanadium(IV) cation $[V^{IV}O(H_2O)_5]^{2+}$ featuring a V=O multiple bond. At higher pH, hydrolysis products $[V^{IV}O(H_2O)_n(OH)_m]^{(2-m)+}$ are formed before precipitation of the hydroxide $V^{IV}O(OH)_2$. The latter is soluble in alkaline solution forming anions $[(V^{IV}O)_2(OH)_5]^{-}$ and $[V^{IV}O(OH)_3]^{-}$. [89] $[V^{IV}O(H_2O)_5]^{2+}$ and $[V^{IV}O(OH)_3]^{-}$ are EPR-active, while the other hydrolysed species and the hydroxide are EPR silent, evidently due to antiferromagnetic spin coupling induced by polymerization. [88]

The coordination geometry of $V^{IV}O^{2+}$ complexes is square pyramidal when

they are penta-coordinated or octahedral, with different degrees of distortion when hexa-coordinated. Regular square pyramidal complexes are observed rarely. When the ligands are bidentate with two different donors, a distortion toward trigonal bipyramidal geometry is detected. The so-called *bare* V^{IV} species are very rare because the oxido ligand of $V^{IV}O^{2+}$ can be lost only under particular conditions, usually leading to the formation of hexa-coordinated complexes of stoichiometry $[V^{IV}L_3]^{2-}$. L^{2-} must be a ligand such as catecholate, 1,2-benzenedithiolate, maleonitriledithiolate or hydroxamate. [88] In these cases, the coordination geometry is intermediate between the octahedron and the trigonal prism, depending on the ligand *bite*.

V^V has no $3d$ electrons and can be studied with ^{51}V nuclear magnetic resonance (NMR) spectroscopy. It exists in alkaline solutions (pH > 13) in the form of vanadate $V^VO_4{}^{3-}$ that can be protonated to form $HV^VO_4{}^{2-}$ and $H_2V^VO_4{}^-$. An increase of concentration and a decrease of pH leads to the formation of oligonuclear condensation products. Di-($V_2O_7{}^{4-}$, $HV_2O_7{}^{3-}$ and $H_2V_2O_7{}^{2-}$), tetra-($V_4O_{12}{}^{4-}$), penta-($V_5O_{15}{}^{5-}$) and deca-vanadates ($V_{10}O_{28}{}^{6-}$, $HV_{10}O_{28}{}^{5-}$, $H_2V_{10}O_{28}{}^{4-}$ and $H_3V_{10}O_{28}{}^{3-}$) are formed depending on the pH, total concentration and ionic strength. Consequently, speciation can be very complicated. VO_4 tetrahedra share corners with neighbouring units in a number of these structures, but the coordination geometry is variable. It can be tetrahedral, as in the case of $V^VO_4{}^{3-}$ and its oligomers, penta-coordinated (with geometries intermediate between square pyramidal and trigonal bipyramidal) or distorted octahedral.

Around pH 7.4, uncomplexed V^V forms only anionic species. The most important are $H_2V^VO_4{}^-$ and $HV^VO_4{}^{2-}$ (for concentrations below 1 μM). On the other hand, the cationic forms V^VO^{3+} and V^VO^{2+} are significantly stabilized by complexation. [88]

Under aerobic conditions and pH = 7, V^{IV} and V^{III} are readily oxidized by O_2 (O_2/H_2O, $E^\circ = 1.23$ V) to V^V. However, in the anaerobic environment of the cytoplasm, specific reducing agents can reduce $V^VO_2{}^+$ to $V^{IV}O^{2+}$, especially if vanadium(V) is in uncomplexed form. The reduction of $H_2V^VO_4{}^-$ is less easy, whereas that of $V^{IV}O^{2+}$ to V^{3+} does not occur under common conditions.

The importance of vanadium as an essential element has been established for a few groups of organism such as *ascidians*, marine polychaete worms, several *Amanita* mushrooms and different bacteria. The function of vanadium, while still elusive in ascidians and polychete, begins to be understood in vanadium-enzyme interaction. [90]

1.2.1 Vanadium in Superior Organisms

Vanadium has been classified as an ultra-trace metal for the superior organisms despite a clear biological role has still to be defined. In humans, nutritional requirement is less than 1 $^{mg}/_{kg}$ diet and it is present in tissues in the range of $^{\mu g}/_{kg}$. However, even if multiple lines of evidence suggest that vanadium is beneficial to human health, its mechanism of action remains obscure. [91, 92]

Experiments carried out on rats and chickens have been suggested the fundamental role of this metal for the organisms health. In chickens, vanadium deficiency caused growth retardation, abnormal skeleton development and delayed feathers of the wings and tail. In rats vanadium deficiency maintained for 4-6 generations caused reduced female fertility, reduced the number of pregnancies per mating, increased mortality of offspring. [93, 94]

Inhibition of kidney tubule (Na^+, K^+)-ATPase by vanadium caused alimentary edema, while increased doses could induce diuresis and natriuresis. [95]

Inclusion of vanadium to the diet increased mineralization of bones and teeth and decreased frequency of caries in rats. [96]

Ammonium vanadate ($20 - 500$ $^{\mu mol}/_{l}$) increased positive inotropic effect on the cat myocardial papillary muscle induced by electric stimulation. [97] However, the inotropic effect of vanadium was dependent on species of the experimental animal, tissue, concentrations and chemical form of vanadium, and also on the presence of other ions and adenylate cyclase activity. Goats fed with less than 10 $^{ng}/_{day}$ diet had difficulty in conceiving and exhibited a high rate of spontaneous abortion. Other experiments with goats demonstrated that a V–deficient diet causes skeletal deformations, convulsion and death within 90 days of birth. In rats, V deprivation resulted in an increase of thyroid weight and a decrease in growth; it also affected the activity of pancreatic amylase and serum lactate dehydrogenase. [98]

These findings indicate that physiological amounts of vanadium affect thyroid hormone and carbohydrate metabolism, providing evidence that it is an essential element for higher forms of life.

Execces in vanadium administration are found to 7 $^{mg}/_{kg}$ and 35 $^{mg}/_{kg}$ respectively for intravenous administration and breathing air. Acute poisoning in animals fed an about tenfold excess of vanadium compounds causes paralysis, convulsion, and eventually death. [90, 99]

1.2.2 Vanadium Enzymes

A whole array of bacteria employ vanadium in several biological functions such as respiratory and electron transport due to the easy interchange between the +V, +IV and +III oxidation states and the capacity to expand its first coordination sphere. [100] For bacteria, vanadate(V), as other high valent metal ions, may be an

electron acceptor in place of O_2, used by higher animals and plants. Furthermore, vanadium bio-catalytic activity is found in the vanadium-dependent Nitrogenases (*VNases*) and haloperoxidases (*VHPOs*) enzymes extracted from marine algae and terrestrial fungi and lichens. The nitrogenases have vanadium incorporated as part of the cofactor, whereas haloperoxidases have vanadate incorporated into the active site. The discovery of these naturally occurring vanadium-containing enzymes is widely quoted in discussions concerning whether or not vanadium is a trace element required for life. [101]

VNases catalyze the reduction of N_2 to NH_4^+, which is a key step in the process of nitrogen fixation. In the natural system, the ATP-driven reduction of N_2 to NH_4^+ (and commonly some hydrazine) is coupled to the reduction of H^+ to H_2: [102]

$$N_2 + 14\,H^+ + 12\,e^- \longrightarrow 2\,NH_4^+ + 3\,H_2 \tag{1.1}$$

Several additional reductive protonation, such as the reduction of CO to alkenes can be catalysed by nitrogenases. [102] The catalytically site, is a metal cluster with the whole core composition $VFe_7[\mu_6{-}C]S_9$ (Figure 1.4 a). [103]

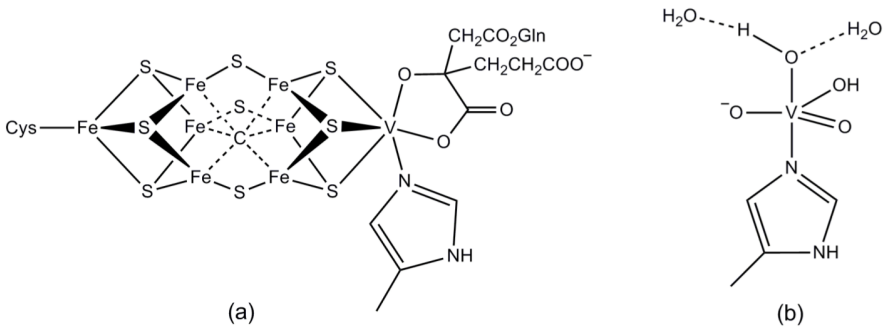

Figure 1.4: The structure of the active centres in (a) *VNases* and (b) *VHPOs*.

VNases can be present (Figure 1.4 a), along with iron-only nitrogenases and the molybdenum nitrogenases, in bacterial strains belonging to the genus *Azotobacter* and in *cyanobacteria* of the genera *Anabaena* and *Nostoc*. The vanadium-based *Azotobacter* nitrogenase is more effective than its molybdenum analogue at low temperatures, and is predominantly expressed when Mo is limited. [100]

Vanadate-dependent haloperoxidases (*VHPOs*) have been found in *cyanobacteria* and in *Streptomyces* bacteria and are directly involved in the utilization

of halides in aquatic environments for catalyze the two-electron oxidative halogenation of organic substrates by halide X^- and hydrogen peroxide as an oxidant: [104–108]

$$X^- + H_2O_2 + H^+ \longrightarrow X^+ + H_2O \tag{1.2}$$

where X^+ is a equilibrium mixture of HOX, X_2 and $X_3{}^-$

If an organic substrate RH capable of electrophilic aromatic substitution is present, the same reaction conditions lead to the substrate halogenations:

$$X^- + H_2O_2 + H^+ + RH \longrightarrow RX + H_2O \tag{1.3}$$

The common substrates for marine *VHPOs* are the halides I^-, Br^- and Cl^- present in seawater. Moreover other substrates such as cyanide, thiocyanate as well as sulfides R_2S are also oxidized:

$$CN^- + H_2O_2 + H^+ \longrightarrow HOCN + H_2O \tag{1.4}$$

$$SCN^- + RH + H_2O_2 + H^+ \longrightarrow SCN{-}R + 2\,H_2O \tag{1.5}$$

$$RR'S + H_2O_2 + H^+ \longrightarrow RR'SO + H_2O \tag{1.6}$$

Despite the vanadate-based catalytic center is non-chiral itself, in the case of prochiral sulfides (RR′S) as substrate, the oxidation occurs enantioselectively suggesting other recognition factors of the active site as responsible of the selectivity.

The *VHPOs*, active site show the $H_2V^VO_4{}^-$ in a trigonal-bipyramid coordination geometry with the axial position occupied by a the εN nitrogen of an histidine side chain and the additional stabilization through a hydrogen bond network with the neighboring amino acid residues (Figure 1.4 b). [105, 109] The catalytic function of V^V is due to its Lewis acidity and its capability to interconvert between a trigonal-bipyramidal and tetragonal-pyramidal arrangement without changes in the oxidation state. [100]

1.2.3 Vanadium in Medicine

The involvement of vanadium in numerous physiological processes and its regulatory effects on the activity of various enzymatic systems suggests its applicability as a therapeutic agent for treatment of various pathologies. Indeed, some vanadium compounds exhibit interesting pharmacological properties. [88, 110, 111] In particular, vanadate(V) and oxidovanadium(IV) derivatives show antidiabetic

activity, [87, 112, 113] growth factor and osteogenic actions, [92, 114–116] anti-tumor properties, [117, 118] cardioprotective and neurological actions, [119–121] antihypertensive, [122] hypocholesterolemic and anti-inflammatory effects. [88]

Insulin Enhancing Activity

The insulin enhancing activity of vanadium compounds is related to the structural, total charge and volume similarity between phosphate and vanadate(V) that make it indistinguishable in the molecular recognition process of phosphate dependent enzymes. [110, 111, 123] In contrast to phosphate ion, for which the pentavalent form is only a transition state achieved on binding to the active site, vanadate(V) forms a stable penta-coordinated compounds, resulting in an inhibition of the respective enzyme. [124–126] An example is the vanadate(V) form of rat prostate acid phosphatase (Figure 1.5). [127]

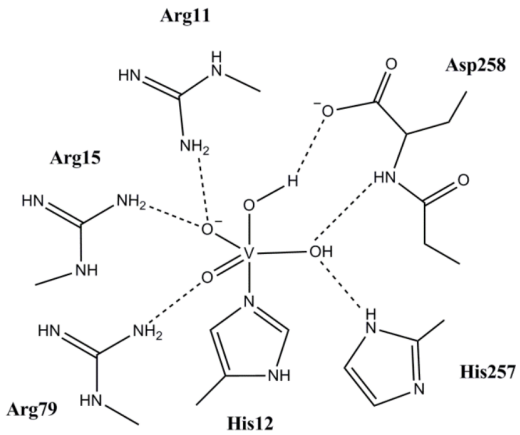

Figure 1.5: Coordination environment of vanadium in the vanadate transition state analog of rat acid phosphatase. Adapted from ref. [104]

The Insulin Receptor (IR), expressed in most of the cellular membrane, is a tetrameric trans-membrane protein, being tyrosine kinases, with two α (IR_α) and two β (IR_β) subunits with the latter exposed to the cytosol. When insulin is bound to IR_α, the tyrosine residues of the intracellular IR_β become phosphorylated triggering the phosphorylation of IR Substrates (IRS), an intracellular signaling protein containing tyrosine residues. The activation of the IRS, in turn starts a signalling cascade, involving the subsequent activation of protein kinases such as phosphatidylinositol 3-kinase (PI3K), protein kinase B (PKB/Akt), and glycogen synthase kinase-3 (GSK-3). [128, 129] PKBs, which bind phosphate through Tyr or

Ser residues, target the glucose transporter GLUT4 stimulating its trans-locations on the cell surface resulting in the glucose intake. [104, 129] In the absence of insulin or in the case of inadequate IR insulin response, the phosphorylation of IR_β subunits and IRS tyrosine residues is counteracted by a protein tyrosine phosphatase (PTP-1B).

On one hand, vanadate(V) and $V^{IV}O^{2+}$ strongly binds the cystein in the active site of PTP-1B inhibiting its phosphatase actions, and for the other hand, vanadate(V) also activate a cytosolic protein tyrosine kinase (cyt-PTK) also related with the signaling path. [90, 130] The synergistic effect of inhibition of phosphatases and activation of kinases lead to the the signal transduction cascade for glucose uptake, (Figure 1.6).

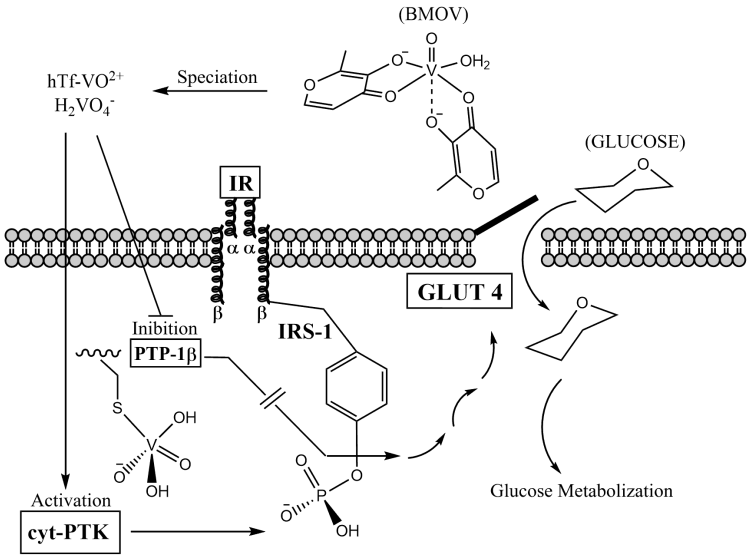

Figure 1.6: Simplified mechanism of action of insulin enhancing vanadium compounds. Adapted from ref. [100]

In addition to the inorganic salts, $V^{IV}O$ coordination compounds have been successfully proposed for antidiabetic treatment. [126] The main advantage lies in the possibility to fine tune the drug properties acting on the chelating functional groups to minimize toxicity, increase stability and absorption in the gastrointestinal tract, control ligand exchange during the bloodstream transport, optimize the cellular uptake modeling the second coordination sphere interactions, and control liability for degradation and re-functionalization within the cell.

$V^{IV}O$ coordination compounds share the same mechanism of action of the

bare V^{V} inorganic salts, in fact in the bilogical environment, ligand exchange, depending on its thermodynamic stability, and subsequent oxidation of V^{IV} to V^{V} complexes can occur and the cellular uptake described above can takes place. Alternative uptake mechanisms have been proposed involving the binary species of transferrin ($V^{IV}O$)hTf and ($V^{V}O_2$)hTf that can be transported to through the cellular membrane via hTf receptor-mediated endocytosis. The hypothesis is supported by a recent study demonstrating that $V^{IV}O^{2+}$ and some of its insulin-enhancing compounds, in the bloodstream, can bind also to holo-hTf. Moreover, $V^{IV}OL_2$ complexes and its oxidation products $V^{V}O_2L_2{}^{-}$ can penetrate the cell membrane via passive diffusion and/or mediated by the formation of neutral ion pairs (Figure 1.6). [90, 130, 131]

The actual challenges are related to i) the development of potential anti-diabetic pro-drug compounds providing, the active species $V^{IV}O^{2+}$, $H_2V^{V}O_4{}^{-}$, and a *free* ligand with a sufficiently long physiological half-life to recombine with, and thus re-solubilise, inorganic vanadium in the process of renal excretion, and ii) deciphering the exact mode of action speciation of vanadium coordination compounds in the presence of blood serum constituents.

A more profound knowledge of the *in vivo* speciation of vanadium compounds should help to understand better the mechanisms of transport, targeting and the mode of action, and thus support the design of potentially efficient vanadium-based drugs. Potential candidates are bioavailable ligands containing the pyrone, the pyridone or the pyridine-carboxylate moiety. A selection of vanadium compounds with insulin-enhancing activity is shown in Figure 1.7. [90]

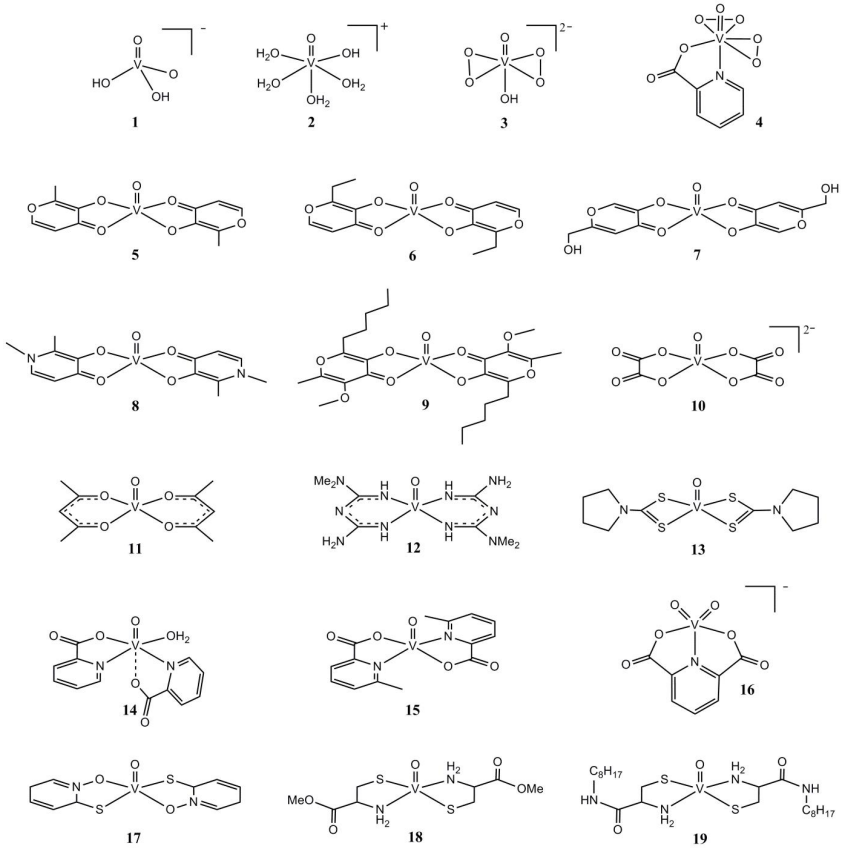

Figure 1.7: Selection of vanadium compounds with insulin-enhancing activity: 1) $H_2VO_4^-$, 2) $[VO(H_2O)_4(OH)]^+$, 3) pervanadate, 4) bis(peroxide)vanadium, 5) bis(maltolato)oxidovanadium(IV) (BMOV) and 6) its ethyl analogue BEOV, 7) bis(kojato)oxidovanadium(IV), 8) the pyridinone complex $[VO(dhp)_2]$, 9) allixinato complex, 10) $[VO(oxalato)_2]$, 11) $[VO(acetylacetonato)_2]$, 12) [VO(N',N'-dimethylbiguanidato$_2$], 13) [VO(pyrrolidine-N-dithiocarbamato$_2$], 14) *cis*-$[VO(picolinato)_2(H_2O)]$, 15) [VO(6-metilpicolinato$_2$], 16) $[VO_2(dipicolinato)]^-$, 17) [VO(pyridin-N-oxide-2-thiolato$_2$], 18) [VO(cysteine-methylester$_2$] and 19) [VO(N-octylcysteinamide$_2$]. Adapted from ref. [90]

Anti-cancer Activity

In recent years the anticancer properties of vanadium compounds have been noticed and several mechanisms have been proposed to explain the inhibition

of cell cycle or the induction of tumor cell death by vanadium derivatives. [132] The main targets for the anti-tumor effects of vanadium are the disruption of cellular metabolism through the generation of ROS (Reactive Oxygen Species such as H_2O_2, O_2^-, $^1O_2$, $HO^{\cdot}$), the alterations of cellular organelles such as lysosomes, mitochondria, the spindle proteins such as actin and tubulin, some signal transduction pathways, cyclins and caspases which in turn play a role in cell cycle arrest and apoptosis. [88] Moreover, cell proliferation can also be disturbed by genotoxic effects of vanadium exerted at the nuclei of the cells and on DNA damage. [104] Important evidence supports the hypothesis that the antitumor action of vanadium compounds is mediated by the inhibition of PTPases. [88] The inhibition of phosphatase can occur by incorporation of vanadate into enzymes that are otherwise activated by phosphate, or indirectly, that is, by the generation of ROS. [118] ROS can deactivate phosphatases (which are over-expressed in several cancer cell types) by oxidation of active site cysteine either directly or via the intermittent formation of ROS. [104, 133] Vanadium compounds can act either as scavengers or generate ROS, depending on the condition, such as pH or coordinated ligands which can cause changing of the redox potential of the complex. It has been demonstrated that vanadate(V) is reduced to oxidovanadium(IV) by glutathione (eq. 1.7) and other potential reducing agents like ascorbate, NADH and $FADH_2$. [104]

$$H_2VO_4^- + GSH + 3\,H^+ \rightarrow VO^{2+} + {}^1/_2GSSG + 3\,H_2O \tag{1.7}$$

$V^{IV}O_2^+$ thus generated can be involved in the production of ROS such as $HO^{\cdot}$ radicals. In equations 1.8-1.12, the reactions related to the mechanism of action of $V^{IV}O$ and V^{V} are reported. Equation 1.11 is a Fenton-like reaction, while equation 1.11 reports the vanadate(V) mediated formation of singlet oxygen from H_2O_2. [134] A complete overview of the anticancer activity of vanadium compounds can be found in refs. [90, 134]

$$V^{IV}O^{2+} + O_2 + e^- \longrightarrow V^{V}O(O_2)^+ \xrightarrow{hydrolisis} V^{V}O^{3+} + H_2O_2 \tag{1.8}$$

$$V^{V}O^{3+} + O_2^{\cdot -} \longrightarrow VO^{2+} + O_2 \tag{1.9}$$

$$V^{V}O^{3+} + O_2^{2-} \longrightarrow VO^{2+} + O_2^{\cdot -} \tag{1.10}$$

$$V^{IV}O^{2+} + H_2O_2^{\cdot} \longrightarrow V^{V}O_2^+ + H^+ + HO^{\cdot} \tag{1.11}$$

$$H_2O_2 + {}^1/_2O_2 \xrightarrow{[H_2VO_4]^-} {}^1O_2 + H_2O \tag{1.12}$$

1.2.4 Speciation and Bloodstream Transport

Despite many vanadium complexes have shown to be highly active against several diseases *in vitro*, the experimental conditions applied (*e.g.* the solvent/incubation medium, pH, *etc.*) may differ considerably from those existing in the living systems. The presence of several endogenous bioligands that may have high affinity for the metal, can partly or fully displace the original ligand(s) during the i) absorption processes in the gastrointestinal tract (in the case of complexes administered orally), ii) transport processes in the blood stream, and iii) reactions with intracellular molecules. [110] The knowledge of the fate of vanadium compounds in the body, the speciation, degradation, rearrangement and the transmembrane transport are fundamental factors in order to understand their mechanism of action. [135, 136] Moreover, a more detailed knowledge of the bio-speciation and pharmacokinetics of prospective metallodrugs may promote the design of more effective compounds in the future. [132]

In this context and regardless of their prospective therapeutic use, one of the most important aspects to understand for metallodrugs is the form and efficiency of transport of the metal ion or complex in the blood stream. The extent and nature of the interaction with blood components has a profound effect on the distribution of the drug in other compartments, and on its therapeutic as well as toxic effects. The knowledge of all these aspects is fundamental for predicting metal interactions with bioligands as well as in drug design and production of new compounds. In the last thirty years, there has been substantial progress on the comprehension of the speciation of vanadium coordination compounds in the blood, especially concerning the well-known insulin-mimetic compounds.

When orally administered, $V^{IV}OL_2$ compounds are subjected to slightly alkaline and oxidizing conditions in the oral cavity, strong acidic in the stomach (pH 2), and slightly alkaline in small intestine. [137] If the ligands are strong enough to avoid oxidation to V^{V} in the oral cavity, at lower pH of the stomach, its ligands may be protonated leading to a partly ligand exchange with other exogenous and endogenous molecules. The quiet distinct pH of the gastrointestinal tract, can induce further modifications of the original pro-drug. So, the compounds must have an adequate hydrophilic-lipophilic balance, be little toxic and sufficiently stable to preserve its original form during the digestion up to reach the bloodstream.

The vanadium compounds present in the blood, depending on their nature and concentration, results distributed between the serum and the erythrocytes, only in +IV and +V oxidation state. The fact is explained with the presence of reducing agents such as ascorbate, catecholamines and cysteines, together with the binding of $V^{IV}O^{2+}$ by bioligands stabilizing the +IV oxidation state. [138]

Table 1.4 lists the relevant bioligands (bL) with their respective concentrations in human plasma.

Among all the low molecular mass bioligands dissolved in serum, experimental

Bioligand	Concentration	Bioligand	Concentration
Albumin	639.4 M	Histidine	77.0 M
Carbonate	24.9M	Immunoglobulin G	101.0 M
Cysteine (Cys-Cys)	10.9 M	Lactate	1.51 μM
Cysteine (Cys)	33.0 M	Oxalate	9.20 M
Citrate	99.0 M	Phosphate	1.10 μM
Glycine	2.30 μM	Sulfate	330 M
Glutamate	60.0 M	Transferrin	29.9 M

[a] Values taken from refs. [139, 140]

Table 1.4: Concentrations of the most important bioligands of blood serum

evidence indicates that only lactate and citrate interact with $V^{IV}O^{2+}$ under physiological conditions, the complexation being favored for the $H_{-1}lact^{2-}$ and $H_{-1}citr^{4-}$ forms (The notation H_{-x} indicates the dissosiation of groups that do not deprotonate in absence of the metal). Therefore, a fundamental role in the bio-speciation of V-drugs in the plasma is covered by proteins. [132, 141]

The interaction of vanadium, in particular in +IV and +V oxidation states, with proteins such as insulin, carboxypeptidase A, carbonic anhydrase B, ferritin, collagen, D-Xylose isomerase, S-adenosylmethionine synthetase, pyruvate kinase, calmodulin, phosphoglucomutase, glucose-6-phosphate dehydrogenase has been reported in the literature. [142] Among the plasma proteins reported in Table 1.4, human serum transferrin (hTf), human serum albumin (HSA) and immunoglobulin G (IgG) seem to be involved in the transport of vanadium and its insulin enhancing compounds. It has been demonstrated that VO^{2+} ion is principally transported in the blood stream as $(VO)_2$hTf and secondly as $(VO)_x$HSA, with $x = 5 - 6$, and $(VO)_x$IgG with $x = 3 - 4$. [142, 143]

The bio-speciation of an insulin-enhancing compound $V^{IV}O(carrier)_2$ after oral administration are represented in Figure 1.8.

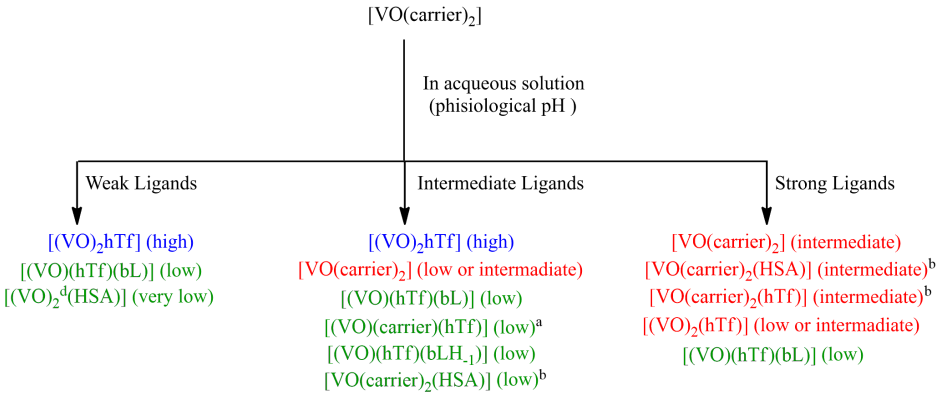

Figure 1.8: Possible biotransformation of a $VO(carrier)_2$ compounds in the blood serum. In brackets the concentration of the several species formed is indicated with high, intermediate, low, and very low. [a] If the carrier is a synergistic anion. [b] If the insulin-enhancing compound assumes the *cis*-octahedral geometry in aqueous solution. [144]

The form in which the vanadium compounds reach the target organs depends not only on the strength of their ligands but also on the geometry assumed by the bis-chelated species in aqueous solution, usually square pyramidal or *cis*-octahedral with a water molecule in the fourth equatorial position, and on the concentration of $V^{IV}O(L)_2$, where L is the carrier ligand. The biospeciation in the blood can be summarized as follows: [145, 146]

i) If the ligand is weak, like 6-methylpicolinate, it cannot compete with the protein for the $V^{IV}O^{2+}$ complexation. [141]

ii) If the carrier ligand has an intermediate strength, like picolinate (pic) or acetylacetonate (acac), the percentage of vanadium bound to the serum proteins is lower. Moreover, if the insulin-enhancing compound can assume a *cis*-octahedral geometry, mixed complexes like *cis*$-VO(L)_2$Protein can be formed, in which an aminoacidc donor such as γ-OAsp/Glu or NHis replace the water molecule in the fourth equatorial position.

iii) Finally, if the carrier is strong, like 1,2-dimethyl-3-hydroxy-4(1*H*)-pyridinone (dhp), the amount of $V(IV)O^{2+}$ ion bound to the protein, as well as the mixed species $V^{IV}O(L)-$Protein decrease significantly favoring the free $V^{IV}O(L)_2$ form. [147] As the total concentration of $V^{IV}O(L)_2$ increases, the relative amount of V^{IV} transported as non-protein bound $V^{IV}O(L)_2$, $V^{IV}O(L)_2$(Protein) or $V^{IV}O(L)_2-$bL species increases.

1.3 Computational Bioinorganic Chemistry

Despite major progresses over the last years, the prediction and elucidation of the exact binding mechanism of metal moieties (being naked or complexed ions) is still a challenging question. Experimental techniques like electrospray ionization (ESI-MS), matrix-assisted laser desorption/ionization time-of-flight (MALDI-ToF-MS), nano electrospray time-of-flight (nESI-Q-ToF-MS) mass spectrometry, are able to accurately determine the metal ion-protein stoichiometry, [148] and in several cases, tandem MS/MS using top-down or bottom-up approaches, can suggest the amino acids involved in binding. [149–151] Spectroscopic techniques as X-ray crystallography, [152] X-ray absorption near edge structure (XANES), extended X-ray absorption fine structure (XAFS), nuclear magnetic resonance (NMR), electron paramagnetic resonance (EPR), electron spin echo envelope modulation (ESEEM), electron nuclear double resonance (ENDOR), circular dichroism (CD) and UV-Vis can provide with three dimensional information of the metal bound structure of a protein but is often inapplicable (*e.g* lability of the drug-protein interaction under crystal conditions and electron beaming, open shell systems *etc.*.) or provide with only partial information on the binding region. [151] Computational methods can represent a valuable complementary approach to experiments and provide with three dimensional models. [23–26] However, a complete simulation of the metal binding process should take into account several variables: i) the wide conformational space to explore on the protein scaffold to detect metal binding site(s) and ii) the intrinsic properties of first coordination sphere of the metal during binding (*e.g.* the directionality of metal-ligand interactions, possible changes in the number of ligands or geometry distortion). The following paragraphs will provide a brief overview of computational techniques useful to address the main steps during the protein-(metal moieties) recognition process and reactivity simulation. An introduction to the theoretical principles behind the techniques can be found in the following chapter.

1.3.1 Molecular Mechanics Approaches

Molecular Mechanics (MM) based approaches are commonly on structural studies of protein systems including metalloproteins. The technique defines the system under analysis basing on classical physics. The potential energy of the whole system is calculated as a function of the nuclear positions using a series of empirical parameters defined in a general force fields. The common parameters are available for well-defined functions for bonded and non-bonded interactions. Among bonded interactions, bonds, angles and dihedral are typically described on the most common force fields (such as AMBER, CHARMM, OPLS *etc.*) [153] using harmonic functions. Non-bonded interactions, like electrostatic and van der Waals (vdW), are generally modeled on the Coulombic framework using

Lennard-Jones potentials. It can be highlighted that the accuracy and the general applicability of the techniques are limited by the number and type of functions that the specific force fields account with. A complete overview of the application of Classical Mechanics to model metal ions can be found in ref. [154]. Despite their high potentiality and relative low computational cost, most of the force fields lack on appropriate parameterization for metal ions, implying a series of limitations in bioinorganics. Several approaches have been developed with the aim to overcome these limitations (Figure 1.9) including: i) representation of the metal ion as simple 12-6 Lennard-Jones (LJ) interacting soft-sphere; [155] ii) bonded models for which the bond between the metal and their ligands are forced during the simulation; [156–159] iii) non-bonded models using dummy atoms. [160–165]

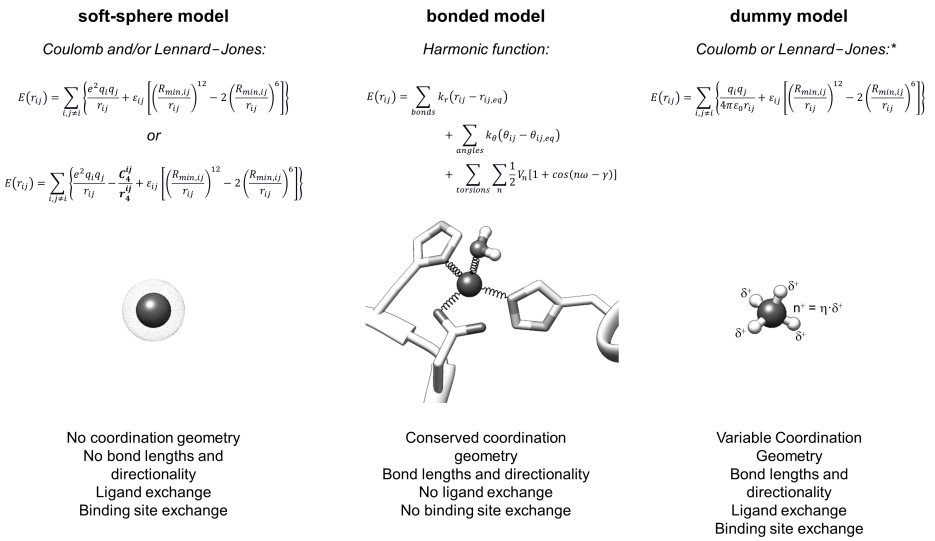

Figure 1.9: Different approaches to deal with metal ions in molecular mechanics framework.

The Lennard-Jones sphere model, describing the ion as a soft-sphere with integer charge, limits the metal interaction with its surrounding of simple vdW or electrostatic interactions. It application was successful applied to main group metals, but shows several limitations for transition metals, for which bond directionality, coordination geometry and metal-donor affinity are fundamental aspects. Moreover, the 12-6 LJ potential neglects metal-induced dipole interaction. Li and Mertz, recently proposed a 12-6-4 LJ-type non-bonded model to account for

charge induced dipole interactions for divalent [166] and highly charged metal ions (up to M^{4+}). [167] The additional empirical C_4 term was added to the standard 12-6 LJ potential and used to represent the ion-induced dipole interaction (Figure 1.9). The developed parameter set is able to reproduce some experimental parameters such as hydration free energy, ion-oxygen distances and coordination number in explicit water solvent with higher accuracy respect the classical LJ potential. Despite transferability on protein biomolecular systems is possible in principle, [166, 167] no specific parameter set accounting for the different donor abilities of the amino acids has been developed so far. On the other hand, bonded models using specific parameters for describing bonds between the metal and its specific donors, give a good description of the coordination geometry, in fact bond distances, angle, dihedral, electrostatic, and vdW terms are included in the force field. However, its application is limited to a very few number of well-defined systems and the parameter portability is often unavailable.[34] Metal center parameters, if not available, could be derived for the specific system with several methods and tools such as FUERZA, [168] MCPB.py [169] or VFFDT [170] and included in standard force fields. The limitation of using strictly *a priori* defined bonds, results in models unable to account for coordination geometry interconversion or ligand exchange, a fundamental aspect in metal-proteins interaction and or in catalysis. [162] Among the several MM based techniques, Molecular Dynamics (MD) must be mentioned as one of the most powerful tools for studying dynamic processes of molecular systems. In particular, in a biomolecular context, MD can account with slight or wide conformational changes even induced for metal binding. The principles of the technique are discussed in details in section 3.4 of the Theoretical Background.

1.3.2 Quantum Mechanics and Hybrid Approaches

In the recent years, quantum mechanics (QM) approaches, in particular those based on Density Functional Theory (DFT), have been increasingly applied to metalloenzymes and metalloproteins in general, becoming a fundamental tool in bionorganic chemistry. The QM approach describes a molecular system at electronic level. [171] DFT and QM methods, in general, allow computing the structure, energy and fine molecular properties with great accuracy and are widely used in inorganic and organometallic systems because of the complexity of the metal properties, in particular of transition metal ions for which subtle electronic effects (*e.g* spin state) must be considered. [172–177] Despite the increasing of the computational power in the last years, QM methods are computationally expensive and applicable only to a reduced number of atoms (about a maximum of 200 atoms). Considering a metalloprotein or a metalloenzyme, this is reflected in the possibility to treat a reduced part of the chemical space, typically the

metal ion, its ligands and few amino acids. A wide used approach to deal with metalloproteins from a QM perspective consists in cutting out the metal containing region generating cluster models without considering the entire size of the system. [178]

A valuable alternative is represented by the so called hybrid Quantum Mechanics/Molecular Mechanics (QM/MM) methods in which part of the system, typically the metal ion, the donors of the first coordination sphere and the neighbor amino acids side chains are treated under QM formalism and the rest of the system at MM level (Figure 1.10). [179–181]

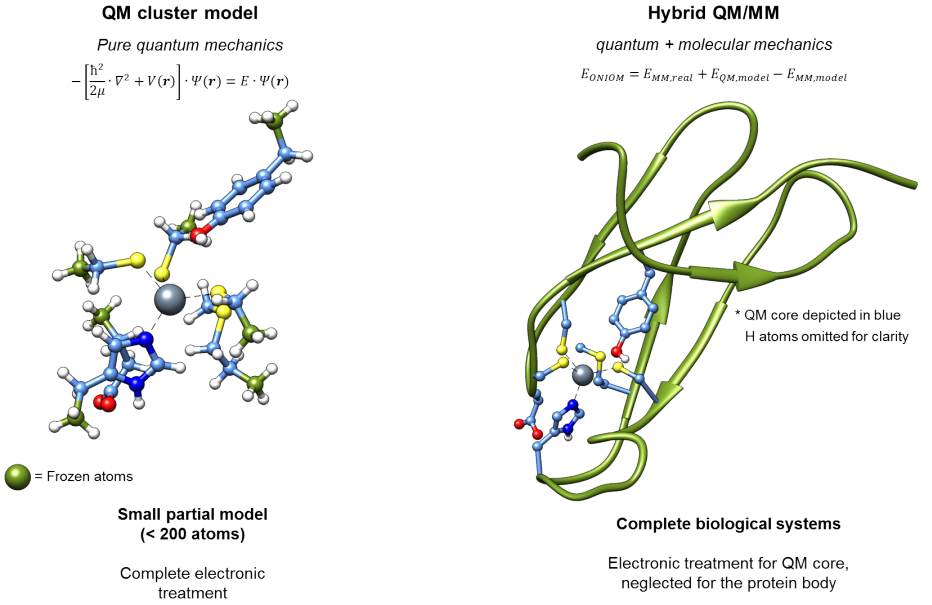

Figure 1.10: Different approaches to deal with metal ions in quantum mechanics hybrid QM/MM frameworks

An important limitation of pure QM or hybrid methodologies described above is that the conformational and chemical spaces that can be explored in a reasonably amount of time are rather limited. To overcome these limitations, QM theory level could be implemented in molecular dynamics via *ab initio* MD (AIMD). The methodology, consists in calculating the forces of each configuration generated along the trajectory using first principles. [182] Two different models are possible: the Born-Oppenheimer (BOMD) and the Car-Parrinello (CPMD) frameworks. In BOMD the forces are extracted optimizing the wavefunction at each step, while on

CPMD the orbitals are treated as classical variables and propagated simultaneously with the particles. [183] Dealing with transition metals, the most convenient and wide used electronic structure method in AIMD simulations is DFT theory level. The treatment of large systems is computationally time demanding, [184] and the hybrid QM/MM/MD approximation is a valuable alternative. [185, 186] The approach, similarly to static QM/MM methodology, consists in partitioning the system extracting on-the-fly the forces of the metal ion and its environment at electronic level, while the rest of the protein and the eventual explicit solvent at MM theory level. [155, 179, 187, 188]

Recently, some efforts to implement mixed QM/MM scoring functions in docking algorithms have been published, [189, 190] and in few cases applied to metalloproteins. [191–193]

Spectroscopy Simulation

During the past decades, many computational methods have been developed to predict the spectroscopic behavior of the chemical species and applied in the context of structural characterization and catalysis. In particular, computation can be used to support the spectroscopic data as well as in the context of structural or reaction simulation to predict the spectroscopic response giving information on the type of transitions expected, and on the role of metal and ligand. [23, 24, 176]

EPR spin Hamiltonian parameters can be computed for most of the transition metal complexes with an odd number of electron and the calculations were based mainly on the DFT methods. [23, 194, 195] The Fermi contact (A^{FC}), the anisotropic hyperfine interaction (A^{D}), and the second-order spin-orbit (SO) coupling (A^{SO}) can be calculated with several packages, for example ORCA. [196] Concerning the prediction of **g** tensor, several approaches are possible even if DFT techniques are the most used, due to the low computational cost, the wide choice of functionals and the possibility to include the SO effect in the calculation of **g**. [194, 195]

UV-vis or electronic absorption spectroscopy is probably the simplest instrumental method to study the electronic transitions of organic and inorganic compounds. For a metal complex, the energy of the transitions could be related to the oxidation states, geometry and electronic structure of a chemical species as well the type of transition, *i.e.* $d-d$ absorptions, metal to ligand charge transfers (MLCT), ligand to metal charge transfers (LMCT), and ligand to ligand charge transfers (LLCT). Time-Dependent density functional theory (TD-DFT) [197, 198] is the most used method to calculate the structure and electronic transitions of metal complexes and such methods describes a given transition as a linear combination of vertical excitations from occupied to virtual molecular orbitals. Consequently, only the dominant character of each transition can be specified.

TD-DFT also provides the electric dipole oscillator strength (f) of the transition from the ground to the excited state related to the transition moment and to the value of ϵ. Although some exceptions have been recently published, up to now only a qualitative agreement between an experimental and calculated spectrum is possible. [199] However, the results can be useful to characterize the the transitions and to describe the electronic structure of a metal complex. The vertical excitations can be obtained within the Linear Response Non-Equilibrium (LRNE), [200] the Linear Response Equilibrium (LRE), [201] the State Specific Non-Equilibrium (SSNE) [202] and State Specific Equilibrium (SSE) [202] frameworks (using a specific combination of functional and basis set).

Vibrational spectroscopy includes several different techniques, among which infrared (IR) and Raman spectroscopy. Both IR and Raman provide information on the vibrations useful for the elucidation of the molecular structure. They are complementary techniques and permit to completely measure the vibrational modes of a molecule. In fact, even if some vibrations may be active in both Raman and IR, these two types of spectroscopy involve different processes and have different selection rules. [203] The software available can calculate the vibrations of the molecules both in the ground and excited states and are able to predict the wavenumber and the intensity of the transitions. The molecular frequencies can be calculated from the second derivative of the energy with respect to the nuclear positions and this is possible for Hartree Fock (HF), post HF and DFT methods. [204]

1.3.3 Molecular Docking

Protein-ligand docking methods were designed for fast and accurate prediction of the binding sites and modes of small organic molecules to proteins and could offer a valuable alternative to MM or MD methodologies. Despite their high performances in terms of computational time and wide system exploration, the method considers the macromolecule as a rigid entity neglecting potential conformational changes in solution or induced for metal binding. The binding energy calculation is based on simplified MM force fields (Scoring Functions, SF) that take into account non-bonded interactions between the ligand (in docking terminology the small molecule) and the protein host. The docking problem is limited to find, generally with a Genetic (GA) or a Montecarlo (MC) algorithm, a series of protein-ligand relative positions (poses), evaluate the energy of each pose throughout the SF and order them in score ranking. Several high quality reviews about Scoring Functions and their performances have been published in the recent years. [205, 206] A SF is generally built as a sum of terms representing differrent interaction type: hydrogen bond (Hbond), electrostatic (elec) or van der Waals (vdW); this latter could be multiplied by an empirical parameter to fine tune its

relative weight. Additionally terms may also incorporated such as desolvation (desolv), interatomic contacts (clashes), ligand torsions (tors), intermolecular ligand interactions or relative to restraints defined (*i.e.* covalent bonds, coordination interactions *etc.*). Unfortunately, commercial docking suites are mainly focused on organic species [207–209] and deal with the coordination bond formation between the metal ion and a donor of an amino acid side chain [111, 210–214] is still an open question in computational bioinorganics. Recent advances in the field will be object of this Thesis work and discussed in detail in the following chapters.

Pattern Recognition Analysis

Pattern recognition algorithms for metal ions binding site could be a valuable alternative to protein-ligand docking methods. Their predictions are generally based on two main approaches: pure sequence or structural analogies between the query apo-protein and well characterized holo-metalloproteins. [215] Neural networks have been also recently introduced for metal ions bindig site prediction, but significant results are still far from being achieved. [216–221]

The most of the available algorithms are limited to selected metals and a reduced set of amino acids are considered as potential donors. Only few methods are devoted to individuate metal binding site without recurring to a template based approach. The recent IonCom server, [222] for example, combines *ab initio* training and template-based algorithms. It focuses on binding site prediction of Zn^{2+}, Cu^{2+}, Fe^{2+}, Fe^{3+}, Ca^{2+}, Mg^{2+}, Mn^{2+}, Na^{+}, K^{+}. The *multisite.py* [223] extension for UCSF Chimera [224] is an automated, structure-based algorithm to locate potential metal binding sites by probing the protein 3D space for regions (blobs, pockets, areas) where two, three or more donor-containing residues featured a β-carbon within a distance threshold of r. The script needs a pdb or mol2 structure as target, a list of donor-containing amino acids and the value for the r threshold. As output, the regions that match with the given criteria are reported as a list of residues and visualized in the protein structure. The assumption on which most of these methods are based is that well characterized metal ions binding sites, defined as specific sequence of amino acids or as 3D regions with defined donors relative positions, could be used to infer the binding residues of a general target protein or peptide. This presents several limitations in prediction capabilities for metals, or specific oxidation states and/or coordination geometries for which scarce templates are available or in the case of metal induced folding.

1.3.4 Multilevel or Integrated Approaches

The metal compound binding site prediction and reactivity can be treated with multiscale strategies combining different level of theory at different stage of the molecular modeling process. Depending on the modeling stage, the accuracy in

energy calculation or the combinational and conformational space to explore could differ. Based on the same concept proposed by Warshel, Karplus and Levitt (Nobel Prize in Chemistry in 2013), multiscale (also known as multilevel, integrative or hierarchical) strategies open major opportunities. Fast and relative inexpensive pattern recognition or docking protocols can be combined with accurate but more time demanding QM or MD techniques (Figure 1.11).

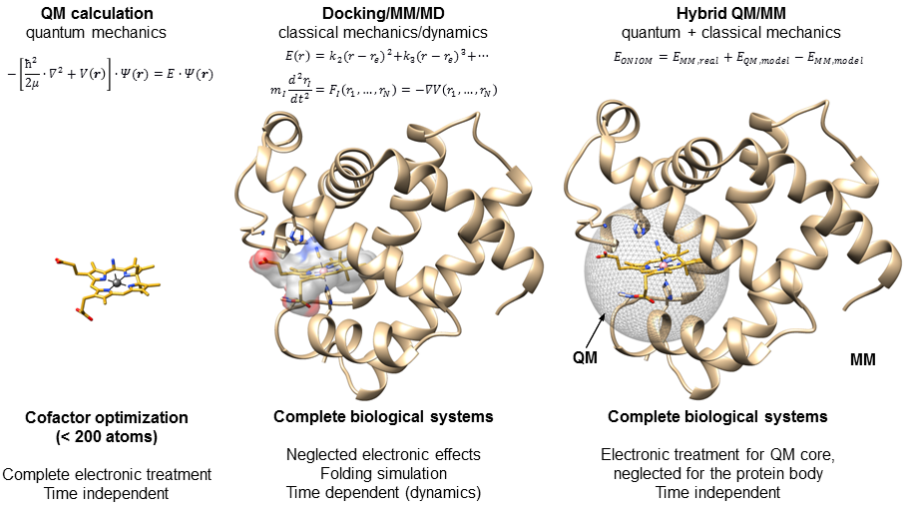

Figure 1.11: The three main families of molecular modeling methods applied in the bioinorganic field. Adapted from ref. [80]

For one side, dockings can be used for a massive exploration of the wide ion-protein relative conformational space. The major drawbacks are: i) the inability to account for slow molecular events like long range motions of the protein scaffold resulting in the loss of information about conformation arrangement upon metal binding; ii) the precision limits of the relative simple scoring function, and iii) the missing of electronic effects. For the other side, MD simulations can account for complete protein flexibility including metal induced fitting of the binding region. Moreover, the explicit solvent effect can be also considered. Finally, QM approaches reach high accuracy and deal with fine electronic phenomena related with first coordination sphere (ligand exchange, spin related or redox processes) and reaction mechanism analysis in the case of metalloenzymes. The combination of the techniques in hierarchical or integrate protocols - where docking gives preliminary information about potential binding sites, MD simulations are employed to explore natural or metal induced conformational changes of the protein

receptor and subsequently QM or QM/MM to optimize, refine the structures of the final adducts, and calculate energies and molecular or spectroscopic properties - results in an interesting and rising approach in molecular modeling landscape. Up to now, full or automated combinations of these levels of theory are still scarce. [225] Interesting examples could be found on metalloenzymes design. [210, 226–232] In a recent work by Mujika and co-workers, through the combination of a recognition algorithm, QM cluster models, MD and QM/MM approaches have been successful applied to identify a series of plausible Al^{3+}-Aβ-peptide models for the aluminum interaction with β-amyloid aggregations involved in Alzheimer's disease. [233]

Bibliography

[1] Zelder F. *Angew. Chem. Int. Ed.*, **2014**, 53(37), 9706–9706.

[2] Medici S., Peana M., Nurchi V. M., Lachowicz J. I., Crisponi G., and Zoroddu M. A. *Coord. Chem. Rev.*, **2015**, 284, 329 – 350.

[3] Mjos K. D. and Orvig C. *Chem. Rev.*, **2014**, 114(8), 4540–4563. PMID: 24456146.

[4] Stillman M. *Angew. Chem.*, **2007**, 119(46), 8895–8896.

[5] Taylor J. P., Hardy J., and Fischbeck K. H. *Science*, **2002**, 296(5575), 1991–1995.

[6] Kelly J. W. *Curr. Opin. Struct. Biol.*, **1998**, 8(1), 101 – 106.

[7] Sigurdsson E. M., Wisniewski T., and Frangione B. *Trends in Mol. Med.*, **2002**, 8(9), 411 – 413.

[8] Chiba T., Hagihara Y., Higurashi T., Hasegawa K., Naiki H., and Goto Y. *J. Biol. Chem.*, **2003**, 278(47), 47016–47024.

[9] Bush A. I. *Trends in Neurosci.*, **2003**, 26(4), 207–214.

[10] Faller P. and Hureau C. *Dalton Trans.*, **2009**, (7), 1080–1094.

[11] Kozlowski H., Janicka-Klos A., Brasun J., Gaggelli E., Valensin D., and Valensin G. *Coord. Chem. Rev.*, **2009**, 253(21), 2665–2685.

[12] Guo Z. and Sadler P. J. *Angew. Chem., Int. Ed.*, **1999**, 38(11), 1512–1531.

[13] Farrell N. *Metal Complexes as Drugs and Chemotherapeutic Agents*, volume 9, pages 809–840. Pergamon, Oxford, 2003.

[14] *Metallotherapeutic Drugs and Metal-Based Diagnostic Agents. The Use of Metals in Medicine.* John Wiley Sons Ltd, Chichester, 2005.

[15] *Interrelations between Essential Metal Ions and Human Diseases.* Springer Science+Business Media, Dordrecht, 2013.

[16] Jones C. and Thornback J. *Medicinal Applications of Coordination Chemistry.* The Royal Society of Chemistry, Cambridge, 2007.

[17] Dabrowiak J. C. *Metals in Medicine.* John Wiley Sons, Ltd, 2009.

[18] *Bioinorganic Medicinal Chemistry.* Wiley-VCH Verlag GmbH Co. KGaA, 2011.

[19] Lewis J. C. *ACS Catal.*, **2013**, 3(12), 2954–2975.

[20] Yu F., Cangelosi V. M., Zastrow M. L., Tegoni M., Plegaria J. S., Tebo A. G., Mocny C. S., Ruckthong L., Qayyum H., and Pecoraro V. L. *Chem. Rev.*, **2014**, 114(7), 3495–3578.

[21] Bos J. and Roelfes G. *Curr. Opin. Struct. Biol.*, **2014**, 19, 135–143.

[22] Pàmies O., Diéguez M., and Bäckvall J.-E. *Adv. Synt. Catal.*, **2015**, 357(8), 1567–1586.

[23] Neese F. *Curr. Opin. Struct. Biol.*, **2003**, 7(1), 125–135.

[24] Neese F. *J. Biol. Inorg. Chem..*, **2006**, 11(6), 702–711.

[25] *Computational Inorganic and BioInorg. Chem.* John Wiley Sons, Ltd, Chichester, 2009.

[26] Farid R., Day T., Friesner R. A., and Pearlstein R. A. *Bioorg. Med. Chem.*, **2006**, 14(9), 3160–3173.

[27] Diéguez M., Bäckvall J.-E., and Pàmies O. *Artificial Metalloenzymes and MetalloDNAzymes in catalysis: from design to applications.* John Wiley & Sons, 2018.

[28] Zoroddu M. A., Aaseth J., Crisponi G., Medici S., Peana M., and Nurchi V. M. *J. Inorg. Biochem.*, **2019**, 195, 120 – 129.

[29] Williams R. J. P. *J. Chem. Soc., Dalton Trans.*, **1991**, pages 539–546.

[30] Da Silva J. F. and Williams R. J. P. *The biological chemistry of the elements: the Inorg. Chem. of life.* Oxford University Press, 2001.

[31] Mertz W. *Science*, **1981**, 213(4514), 1332–1338. cited By 614.

[32] Horovitz C. *J. Trace Elem. Med. Biol.*, **1988**, 2(3), 135–144. cited By 17.

[33] Maret W. *Int. J. Mol. Sci.*, **2016**, 17(1), 66.

[34] Bertini G., Gray H., Gray H. B., Stiefel E., Valentine J. S., and Stiefel E. I. *Biological Inorg. Chem.: structure and reactivity.* University Science Books, 2007.

[35] Kaim W., Schwederski B., and Klein A. *BioInorg. Chem.–Inorganic Elements in the Chemistry of Life: An Introduction and Guide.* John Wiley & Sons, 2013.

[36] Pearson R. G. *J. Am. Chem. Soc.*, **1963**, 85(22), 3533–3539.

[37] Hughes M. and Wiley J. *FEBS Lett.*, **1981**.

[38] Lakatos B., Szentmihályi K., Vinkler P., Balla J., and Balla G. *Orv. Hetil.*, **2004**, 145(25), 1315–1319. cited By 4.

[39] Ronconi L. and Sadler P. J. *Coord. Chem. Rev.*, **2007**, 251(13), 1633 – 1648. 37th International Conference on Coordination Chemistry, Cape Town, South Africa.

[40] Sadler P. J. Inorg. chem. and drug design. In *Adv. Inorg. Chem.*, volume 36, pages 1–48. Elsevier, 1991.

[41] Guo Z. and Sadler P. *Academic Press: San Diego, CA, USA*, **2000**, 49, 183–306.

[42] Orvig C. and Abrams M. J. *Chem. Rev.*, **1999**, 99(9), 2201–2204. PMID: 11749478.

[43] Guo Z. and Sadler P. J. *Angew. Chem. Int. Ed.*, **1999**, 38(11), 1512–1531.

[44] Alessio E. *Bioinorganic medicinal chemistry*. John Wiley & Sons, 2011.

[45] Mukherjee A. and Sadler P. J. *Metals in Medicine: Therapeutic Agents*, pages 1–47. American Cancer Society, 2009.

[46] Bruijnincx P. C. and Sadler P. J. *Curr. Opin. Struct. Biol.*, **2008**, 12(2), 197 – 206. Biocatalysis and biotransformation/BioInorg. Chem.

[47] Desoize B. *Anticancer Res.*, **2004**, 24(3 A), 1529–1544. cited By 202.

[48] Pizarro A. M. and Sadler P. J. *Biochimie*, **2009**, 91(10), 1198 – 1211. Metals in health.

[49] Orvig C., Thompson K. H., Battell M., and McNeill J. H. *Vanadium compounds as insulin mimics*, volume 31 of *Metal ions in Biological Systems*, page 575–594. Marcel Dekker Inc., New York, 1995.

[50] Thompson K. H., Lichter J., LeBel C., Scaife M. C., McNeill J. H., and Orvig C. *J. Inorg. Biochem.*, **2009**, 103(4), 554–558.

[51] Gielen M. and Tiekink E. R. *Metallotherapeutic drugs and metal-based diagnostic agents: the use of metals in medicine*. John Wiley & Sons, 2005.

[52] Finkelstein A., Walz D., Batista V., Mizraji M., Roisman F., and Misher A. *Ann. Rheum. Dis.*, **1976**, 35(3), 251–257.

[53] Dean T. C., Yang M., Liu M., Grayson J. M., DeMartino A. W., Day C. S., Lee J., Furdui C. M., and Bierbach U. *ACS Med. Chem. Lett.*, **2017**, 8(5), 572–576.

[54] Mirabell C. K., Johnson R. K., Hill D. T., Faucette L. F., Girard G. R., Kuo G. Y., Sung C. M., and Crooke S. T. *J. Med. Chem.*, **1986**, 29(2), 218–223.

[55] Reedijk J. and Lohman P. *Pharm. Weekblad, Sci. Ed.*, **October 1985**, 7(5), 173—180.

[56] Ahmad S., Isab A. A., Ali S., and Al-Arfaj A. R. *Polyhedron*, **2006**, 25(7), 1633 – 1645.

[57] Jamieson E. R. and Lippard S. J. *Chem. Rev.*, **1999**, 99(9), 2467–2498. PMID: 11749487.

[58] Wang D. and Lippard S. J. *Nat. Rev. Drug Discov.*, **2005**, 4(4), 307.

[59] Kelland L. *Nat. Rev. Cancer*, **2007**, 7(8), 573.

[60] Graham J., Muhsin M., and Kirkpatrick P. Oxaliplatin, 2004.

[61] Chaney S. G., Campbell S. L., Bassett E., and Wu Y. *Crit. Rev. Oncol. Hematol.*, **2005**, 53(1), 3 – 11.

[62] Leijen S., Burgers S. A., Baas P., Pluim D., Tibben M., Werkhoven E., van, Alessio E., Sava G., Beijnen J. H., and Schellens J. H. M. *Invest. New Drugs*, **Feb 2015**, 33(1), 201–214.

[63] Hartinger C. G., Zorbas-Seifried S., Jakupec M. A., Kynast B., Zorbas H., and Keppler B. K. *J. Inorg. Biochem.*, **2006**, 100(5), 891 – 904. Special Issue Containing Contributions from the 12th International Conference on Biological Inorg. Chem.

[64] Sava G., Bergamo A., Zorzet S., Gava B., Casarsa C., Cocchietto M., Furlani A., Scarcia V., Serli B., Iengo E., and others . *Eur. J. Cancer*, **2002**, 38(3), 427–435.

[65] Dyson P. J. and Sava G. *Dalton Trans.*, **2006**, (16), 1929–1933.

[66] Scolaro C., Bergamo A., Brescacin L., Delfino R., Cocchietto M., Laurenczy G., Geldbach T. J., Sava G., and Dyson P. J. *J. Med. Chem.*, **2005**, 48(12), 4161–4171.

[67] Bertini I., Ciurli S., and Luchinat C. *The electronic structure of FeS centers in proteins and models a contribution to the understanding of their electron transfer properties*, pages 1–53. Springer Berlin Heidelberg, Berlin, Heidelberg, 1995.

[68] Nordlund P. and Eklund H. *Curr. Opin. Struct. Biol.*, **1995**, 5(6), 758 – 766.

[69] Que L. and Ho R. Y. N. *Chem. Rev.*, **1996**, 96(7), 2607–2624. PMID: 11848838.

[70] Adman E. T. Copper protein structures. In Anfinsen C., Edsall J. T., Richards F. M., and Eisenberg D. S., editors, *Metalloproteins: Structural Aspects*, volume 42 of *Advances in Protein Chemistry*, pages 145 – 197. Academic Press, 1991.

[71] Lipscomb W. N. and Sträter N. *Chem. Rev.*, **1996**, 96(7), 2375–2434. PMID: 11848831.

[72] McCall K. A., Huang C.-c., and Fierke C. A. *J. Nutr.*, **2000**, 130(5), 1437S–1446S.

[73] Ermler U., Grabarse W., Shima S., Goubeaud M., and Thauer R. K. *Curr. Opin. Struct. Biol.*, **1998**, 8(6), 749 – 758.

[74] Wieghardt K. *Angew. Chem. Int. Ed. in English*, **1994**, 33(7), 725–728.

[75] Yocum C. F. and Pecoraro V. L. *Curr. Opin. Struct. Biol.*, **1999**, 3(2), 182 – 187.

[76] Carrell T., Tyryshkin A., and Dismukes G. *J. Biol. Inorg. Chem.*., **2002**, 7(1), 2–22.

[77] Hille R. *Trends Biochem. Sci.*, **2002**, 27(7), 360–367.

[78] Williams R. and Da Silva J. F. *Biochem. Biophys. Res. Commun.*, **2002**, 292(2), 293–299.

[79] Kobayashi M. and Shimizu S. *Eur. J. Biochem.*, **1999**, 261(1), 1–9.

[80] Rodríguez-Guerra J., Alonso-Cotchico L., Sciortino G., Lledós A., and Maréchal J.-D. *Artificial Metalloenzymes and MetalloDNAzymes in Catalysis: From Design to Applications*, **2018**.

[81] Yu F., Cangelosi V. M., Zastrow M. L., Tegoni M., Plegaria J. S., Tebo A. G., Mocny C. S., Ruckthong L., Qayyum H., and Pecoraro V. L. *Chem. Rev.*, **2014**, 114(7), 3495–3578. PMID: 24661096.

[82] Wilson M. E. and Whitesides G. M. *J. Am. Chem. Soc.*, **1978**, 100(1), 306–307.

[83] Heinisch T. and Ward T. R. *Curr. Opin. Struct. Biol.*, **2010**, 14(2), 184 – 199. Biocatalysis and Biotransformation/BioInorg. Chem.

[84] Prier C. K. and Arnold F. H. *J. Am. Chem. Soc.*, **2015**, 137(44), 13992–14006. PMID: 26502343.

[85] Chatterjee A., Mallin H., Klehr J., Vallapurackal J., Finke A. D., Vera L., Marsh M., and Ward T. R. *Chem. Sci.*, **2016**, 7(1), 673–677.

[86] Dürrenberger M., Heinisch T., Wilson Y. M., Rossel T., Nogueira E., Knörr L., Mutschler A., Kersten K., Zimbron M. J., Pierron J., and others . *Angew. Chem. Int. Ed.*, **2011**, 50(13), 3026–3029.

[87] Thompson K. H. and Orvig C. *J. Chem. Soc., Dalton Trans.: Inorg. Chem.*, **2000**, (17), 2885–2892.

[88] Barrio D. and Etcheverry S. *Curr. Med. Chem.*, **2010**, 17(31), 3632–3642. cited By 34.

[89] Boas L. V. and Costa Pessoa J. **1987**.

[90] Rehder D. *Bioinorganic vanadium chemistry*, volume 30. John Wiley & Sons, 2008.

[91] Nielsen F. H. New essential trace elements for the life sciences. In *Nuclear Analytical Methods in the Life Sciences*, pages 599–611. Springer, 1990.

[92] Nielsen F. H. *Vanadium in mammalian physiology and nutrition*, volume 31 of *Vanadium and its role in life*, pages 543–574. Marcel Dekker, New York, 1995.

[93] Hopkins L. and Mohr H. Effect of vanadium deficiency on plasma cholesterol of chicks. In *Fed. Proc.*, volume 30, page A462. Fed. Amer. Soc. Exp. Biol. 9650 ROCKVILLE PIKE, BETHESDA, MD 20814-3998, 1971.

[94] Golden M. H. and Golden B. E. *Br. Med. Bull.*, **1981**, 37(1), 31–6.

[95] Söremark R., Ullberg S., and Appelgren L.-E. *Acta Odontol. Scand.*, **1962**, 20(3), 225–232.

[96] Akera T., Temma K., and Takeda K. Cardiac actions of vanadium. In *Fed. Proc.*, volume 42, pages 2984–2988, 1983.

[97] Nechay B. R. *Annual Rev. Pharmacol. Toxicol.*, **1984**, 24(1), 501–524.

[98] Nielsen F. H. *The Nutritional Essentiality and Physiological Metabolism of Vanadium in Higher Animals*, chapter 23, pages 297–307.

[99] Priestley J. *Philos. Trans. R. Soc. Lond.*, **1876**, (166), 495–556.

[100] Rehder D. *Metallomics*, **2015**, 7(5), 730–742.

[101] Tracey A. S., Willsky G. R., and Takeuchi E. S. *Vanadium: chemistry, biochemistry, pharmacology and practical applications.* CRC press, 2007.

[102] Hu Y., Lee C. C., and Ribbe M. W. *Science*, **2011**, 333(6043), 753–755.

[103] Fay A. W., Blank M. A., Lee C. C., Hu Y., Hodgson K. O., Hedman B., and Ribbe M. W. *J. Am. Chem. Soc.*, **2010**, 132(36), 12612–12618.

[104] Rehder D. *Future Med. Chem.*, **2012**, 4(14), 1823–1837.

[105] Messerschmidt A. and Wever R. *Proc. Natl. Acad. Sci.*, **1996**, 93(1), 392–396.

[106] Winter J. M. and Moore B. S. *J. Biol. Chem.*, **2009**, 284(28), 18577–18581.

[107] Kaysser L., Bernhardt P., Nam S.-J., Loesgen S., Ruby J. G., Skewes-Cox P., Jensen P. R., Fenical W., and Moore B. S. *J. Am. Chem. Soc.*, **2012**, 134(29), 11988–11991.

[108] Bernroitner M., Zamocky M., Furtmüller P. G., Peschek G. A., and Obinger C. *J. Exp. Bot.*, **2009**, 60(2), 423–440.

[109] Zampella G., Fantucci P., Pecoraro V. L., and De Gioia L. *J. Am. Chem. Soc.*, **2005**, 127(3), 953–960.

[110] Pessoa J. C., Etcheverry S., and Gambino D. *Coord. Chem. Rev.*, **2015**, 301, 24–48.

[111] Pessoa J. C., Garribba E., Santos M. F., and Santos-Silva T. *Coord. Chem. Rev.*, **2015**, 301, 49–86.

[112] Sakurai H., Yasui H., and Adachi Y. *Expert Opin. Investig. Drugs*, **2003**, 12(7), 1189–1203.

[113] Thompson K. H. and Orvig C. *J. Inorg. Biochem.*, **2006**, 100(12), 1925–1935.

[114] Srivastava A. and Mehdi M. *Diabet. Med.*, **2005**, 22(1), 2–13.

[115] Cortizo A. M. and Etcheverry S. B. *Mol. Cell. Biochem.*, **1995**, 145(2), 97–102.

[116] Barrio D. and Etcheverry S. *Can. J. Physiol. Pharmacol.*, **2006**, 84(7), 677–686.

[117] Djordjevic C. and Wampler G. L. *J. Inorg. Biochem.*, **1985**, 25(1), 51–55.

[118] Evangelou A. M. *Crit. Rev. Oncol. Hematol.*, **2002**, 42(3), 249–265.

[119] Kris-Etherton P. M., Hecker K. D., Bonanome A., Coval S. M., Binkoski A. E., Hilpert K. F., Griel A. E., and Etherton T. D. *Am. J. Med.*, **2002**, 113(9), 71–88.

[120] Bhuiyan M. S. and Fukunaga K. *J. Pharmacol. Sci.*, **2009**, 110(1), 1–13.

[121] Shioda N., Han F., and Fukunaga K. *Int. Rev. Neurobiol.*, **2009**, 85, 375–387.

[122] Bhanot S., Michoulas A., and McNeill J. H. *Mol. Cell. Biochem.*, **1995**, 153(1-2), 205–209.

[123] McLauchlan C. C., Peters B. J., Willsky G. R., and Crans D. C. *Coord. Chem. Rev.*, **2015**, 301, 163–199.

[124] Shechter Y. and Karlish S. J. D. *Nature*, **1980**, 284(5756), 556–558.

[125] Pederson R. A., Ramanadham S., Buchan A. M. J., and McNeill J. H. *Diabetes*, **1989**, 38(11), 1390–1395.

[126] Orvig C., Caravan P., Gelmini L., Glover N., Herring F. G., Li H., McNeill J. H., Rettig S. J., and Setyawati I. A. *J. Am. Chem. Soc.*, **1995**, 117(51), 12759–12770.

[127] Lindqvist Y., Schneider G., and Vihko P. *Eur. J. Biochem.*, **1994**, 221(1), 139–142.

[128] Hiromura M., Adachi Y., Machida M., Hattori M., and Sakurai H. *Metallomics*, **2009**, 1(1), 92–100.

[129] Sakurai H., Yoshikawa Y., and Yasui H. *Chem. Soc. Rev.*, **2008**, 37(11), 2383–2392.

[130] Wei Y., Zhang C., Zhao P., Yang X., and Wang K. *J. Inorg. Biochem.*, **2011**, 105(8), 1081–1085.

[131] Sanna D., Palomba J., Lubinu G., Buglyo P., Nagy S., Perdih F., and Garribba E. *J. Med. Chem.*, **2018**, 62(2), 654–664.

[132] Costa Pessoa J. and Tomaz I. *Curr. Med. Chem.*, **2010**, 17(31), 3701–3738.

[133] Liu T.-T., Liu Y.-J., Wang Q., Yang X.-G., and Wang K. *J. Biol. Inorg. Chem..*, **2012**, 17(2), 311–320.

[134] Rehder D. *Influence of Vanadium Compounds on Cellular Functions*, pages 157–201. John Wiley & Sons, Ltd, 2008.

[135] Yoshikawa Y., Sakurai H., Crans D. C., Micera G., and Garribba E. *Dalton Trans.*, **2014**, 43(19), 6965–6972.

[136] Gorzsás A., Andersson I., and Pettersson L. *Eur. J. Inorg. Chem.*, **2006**, 2006(18), 3559–3565.

[137] Kiss T., Jakusch T., Hollender D., Dörnyei , Enyedy A., Pessoa J. C., Sakurai H., and Sanz-Medel A. *Coord. Chem. Rev.*, **2008**, 252(10–11), 1153–1162.

[138] Harris W. R., Friedman S. B., and Silberman D. *J. Inorg. Biochem.*, **1984**, 20(2), 157–169.

[139] Harris W. R. *Clin. Chem.*, **1992**, 38(9), 1809–1818.

[140] Schaller J., Gerber S., Kaempfer U., Lejon S., and Trachsel C. *Human blood plasma proteins: structure and function.* John Wiley & Sons, 2008.

[141] Sanna D., Micera G., and Garribba E. *Inorg. Chem.*, **2009**, 48(13), 5747–5757.

[142] Chasteen N. D. *Vanadium proteins*, volume 31 of *Metal ions in biological systems*, book section 7, pages 231–248. Marcel Dekker, New York, 1995.

[143] Sanna D., Bíró L., Buglyó P., Micera G., and Garribba E. *J. Inorg. Biochem.*, **2012**, 115(0), 87–99.

[144] Sanna D., Micera G., and Garribba E. *Inorg. Chem.*, **2010**, 49(1), 174–187. PMID: 19947643.

[145] Costa Pessoa J. and Tomaz I. *Curr. Med. Chem.*, **2010**, 17(31), 3701–3738.

[146] Sanna D., Buglyó P., Micera G., and Garribba E. *J. Biol. Inorg. Chem..*, **2010**, 15(6), 825–839.

[147] Kiss E., Kawabe K., Tamura A., Jakusch T., Sakurai H., and Kiss T. *J. Inorg. Biochem.*, **2003**, 95(2-3), 69–76.

[148] Naylor S. and Kumar R. *Emerging role of mass spectrometry in structural and functional proteomics*, volume 65, pages 217–248. Elsevier, 2003.

[149] Møller C., Sprenger R. R., Stürup S., and Højrup P. *Anal Bioanal. Chem.*, **2011**, 401(5), 1619.

[150] Wenzel M. and Casini A. *Coord. Chem. Rev.*, **2017**, 352, 432–460.

[151] Shi W. and Chance M. R. *Curr. Opin. Struct. Biol.*, **2011**, 15(1), 144–148.

[152] Ward J., Ollmann E., Maxey E., and Finney L. A. *X-Ray Absorption Spectroscopy of Metalloproteins*, pages 171–187. Springer, 2014.

[153] Ponder J. W. and Case D. A. *Force fields for protein simulations*, volume 66, pages 27–85. Elsevier, 2003.

[154] Li P. and Merz K. M. *Chem. Rev.*, **2017**, 117(3), 1564–1686.

[155] Li P., Roberts B. P., Chakravorty D. K., and Merz J., Kenneth M. *J. Chem. Theory Comput.*, **2013**, 9(6), 2733–2748.

[156] Hancock R. D. *Prog. Inorg. Chem.*, **1989**, 37, 187291.

[157] Hancock R. D. *Acc. Chem. Res.*, **1990**, 23(8), 253–257.

[158] Lin F. and Wang R. *J. Chem. Theory Comput.*, **2010**, 6(6), 1852–1870.

[159] Peters M. B., Yang Y., Wang B., Füsti-Molnár L., Weaver M. N., and Merz K. M. *J. Chem. Theory Comput.*, **2010**, 6(9), 2935–2947.

[160] Pang Y.-P. *J. Mol. Model.*, **1999**, 5(10), 196–202.

[161] Oelschlaeger P., Klahn M., Beard W. A., Wilson S. H., and Warshel A. *J. Mol. Biol.*, **2007**, 366(2), 687–701.

[162] Saxena A. and Sept D. *J. Chem. Theory Comput.*, **2013**, 9(8), 3538–3542.

[163] Duarte F., Bauer P., Barrozo A., Amrein B. A., Purg M., Åqvist J., and Kamerlin S. C. L. *J. Phys. Chem. B*, **2014**, 118(16), 4351–4362.

[164] Liao Q., Kamerlin S. C. L., and Strodel B. *J. Phys. Chem. Lett.*, **2015**, 6(13), 2657–2662.

[165] Liao Q., Pabis A., Strodel B., and Kamerlin S. C. L. *J. Phys. Chem. Lett.*, **2017**, 8(21), 5408–5414.

[166] Li P. and Merz J., Kenneth M. *J. Chem. Theory Comput.*, **2014**, 10(1), 289–297.

[167] Li P., Song L. F., and Merz K. M. *J. Phys. Chem. B*, **2015**, 119(3), 883–895.

[168] Seminario J. M. *Int. J. Quantum Chem.*, **1996**, 60(7), 1271–1277.

[169] Li P. and Merz K. M. *J. Chem. Inf. Model.*, **2016**, 56(4), 599–604.

[170] Zheng S., Tang Q., He J., Du S., Xu S., Wang C., Xu Y., and Lin F. *J. Chem. Inf. Model.*, **2016**, 56(4), 811–818.

[171] Hohenberg P. and Kohn W. *Phys. Rev.*, **1964**, 136(3B), B864–B871.

[172] Sameera W. M. C. and Maseras F. *Wiley Interdiscip. Rev. Comput. Mol.*, **2012**, 2(3), 375–385.

[173] Sperger T., Sanhueza I. A., Kalvet I., and Schoenebeck F. *Chem. Rev.*, **2015**, 115(17), 9532–9586.

[174] Balcells D., Clot E., Eisenstein O., Nova A., and Perrin L. *Acc. Chem. Res.*, **2016**, 49(5), 1070–1078.

[175] Cramer C. J. and Truhlar D. G. *Phys. Chem. Chem. Phys.*, **2009**, 11(46), 10757–10816.

[176] Neese F. *Coord. Chem. Rev.*, **2009**, 253(5), 526–563.

[177] Becke A. D. *J. Chem. Phys.*, **2014**, 140(18), 18A301.

[178] Siegbahn P. E. and Himo F. *Wiley Interdiscip. Rev. Comput. Mol.*, **2011**, 1(3), 323–336.

[179] Senn H. M. and Thiel W. *Angew. Chem. Int. Ed.*, **2009**, 48(7), 1198–1229.

[180] Chung L. W., Sameera W. M. C., Ramozzi R., Page A. J., Hatanaka M., Petrova G. P., Harris T. V., Li X., Ke Z., Liu F., Li H.-B., Ding L., and Morokuma K. *Chem. Rev.*, **2015**, 115(12), 5678–5796.

[181] Vidossich P., Lledós A., and Ujaque G. *Realistic Simulation of Organometallic Reactivity in Solution by Means of First-Principles Molecular Dynamics*, pages 81–106. Springer, 2015.

[182] Marx D. and Hutter J. *Ab initio molecular dynamics: basic theory and advanced methods*. Cambridge University Press, 2009.

[183] Car R. and Parrinello M. *Phys. Rev. Letters*, **1985**, 55(22), 2471–2474.

[184] Sulpizi M., Raugei S., VandeVondele J., Carloni P., and Sprik M. *J. Phys. Chem. B*, **2007**, 111(15), 3969–3976.

[185] Warshel A. and Levitt M. *J. Mol. Biol.*, **1976**, 103(2), 227–249.

[186] Singh U. C. and Kollman P. A. *J. Comput. Chem.*, **1986**, 7(6), 718–730.

[187] Vidossich P., Lledós A., and Ujaque G. *Acc. Chem. Res.*, **2016**, 49(6), 1271–1278.

[188] Vidossich P. and Magistrato A. *Biomolecules*, **2014**, 4(3), 616.

[189] Raha K. and Merz K. M. *J. Med. Chem.*, **2005**, 48(14), 4558–4575.

[190] Fong P., McNamara J. P., Hillier I. H., and Bryce R. A. *J. Chem. Inf. Model.*, **2009**, 49(4), 913–924.

[191] Raha K. and Merz K. M. *J. Am. Chem. Soc.*, **2004**, 126(4), 1020–1021.

[192] Chaskar P., Zoete V., and Röhrig U. F. *J. Chem. Inf. Model.*, **2014**, 54(11), 3137–3152.

[193] Hayik S. A., Dunbrack R., and Merz K. M. *J. Chem. Theory Comput.*, **2010**, 6(10), 3079–3091.

[194] Sciortino G., Lubinu G., Maréchal J.-D., and Garribba E. *Magnetochemistry*, **2018**, 4(4), 55.

[195] Sanna D., Sciortino G., Ugone V., Micera G., and Garribba E. *Inorg. Chem.*, **2016**, 55(15), 7373–7387.

[196] Neese F. O. *ORCA - An Ab Initio, DFT and Semiempirical Program Package, Version 3.0.* Max-Planck-Institute for Chemical Energy Conversion: Mülheim a. d. Ruhr, 2013.

[197] Casida M. E. *J. Mol. Struc-THEOCHEM*, **2009**, 914(1-3), 3–18.

[198] Runge E. and Gross E. K. *Phys. Rev. Letters*, **1984**, 52(12), 997.

[199] Sciortino G., Lihi N., Czine T., Maréchal J.-D., Lledós A., and Garribba E. *Int. J. Quantum Chem.*, **2018**, 118(16), e25655.

[200] Cammi R., Mennucci B., and Tomasi J. *J. Phys. Chem. A*, **2000**, 104(23), 5631–5637.

[201] Tomasi J., Mennucci B., and Cammi R. *Chem. Rev.*, **2005**, 105(8), 2999–3094.

[202] Cammi R., Corni S., Mennucci B., and Tomasi J. *J. Chem. Phys.*, **2005**, 122(10), 104513.

[203] Sridharan K. *Spectral methods in transition metal complexes.* Elsevier, 2016.

[204] Koch W. and Holthausen M. A chemist's guide to density functional theory. wileyvch, verlag gmbh, 2000.

[205] Liu J. and Wang R. **2015**, 55(3), 475–482.

[206] Huang S.-Y., Grinter S. Z., and Zou X. *Phys. Chem. Chem. Phys.*, **2010**, 12(40), 12899–12908.

[207] Yuriev E., Holien J., and Ramsland P. A. *J. Mol. Recognit.*, **2015**, 28(10), 581–604.

[208] Warren G. L., Andrews C. W., Capelli A.-M., Clarke B., LaLonde J., Lambert M. H., Lindvall M., Nevins N., Semus S. F., Senger S., and others . *J. Med. Chem.*, **2006**, 49(20), 5912–5931.

[209] Guedes I. A., Magalhães C. S., de, and Dardenne L. E. *Biophys. Rev.*, **2014**, 6(1), 75–87.

[210] Munoz Robles V., Ortega-Carrasco E., Alonso-Cotchico L., Rodriguez-Guerra J., Lledos A., and Marechal J.-D. *ACS Catal.*, **2015**, 5(4), 2469–2480.

[211] Timerbaev A. R., Hartinger C. G., Aleksenko S. S., and Keppler B. K. *Chem. Rev.*, **2006**, 106(6), 2224–2248.

[212] Vincent J. B. and Love S. *Biochim. Biophys. Acta, Gen. Subj.*, **2012**, 1820(3), 362–378.

[213] Casini A. and Reedijk J. *Chem. Sci.*, **2012**, 3(11), 3135–3144.

[214] Sanna D., Micera G., and Garribba E. *Inorg. Chem.*, **2013**, 52(20), 11975–11985.

[215] Akcapinar G. B. and Sezerman O. U. *Biosci. Rep.*, **2017**, 37(2), BSR20160179.

[216] Brylinski M. and Skolnick J. *Proteins: Struct., Funct., Bioinf.*, **2011**, 79(3), 735–751.

[217] Liu T. and Altman R. B. *BMC Struct. Biol.*, **2009**, 9(1), 72.

[218] Passerini A., Andreini C., Menchetti S., Rosato A., and Frasconi P. *BMC bioinformatics*, **2007**, 8(1), 39.

[219] Levy R., Edelman M., and Sobolev V. *Proteins: Struct., Funct., Bioinf.*, **2009**, 76(2), 365–374.

[220] Cai C., Han L., Ji Z. L., Chen X., and Chen Y. Z. *Nucleic Acids Res.*, **2003**, 31(13), 3692–3697.

[221] Haberal İ. and Oğul H. *Mol. Inform.*, **2019**.

[222] Hu X., Dong Q., Yang J., and Zhang Y. *Bioinformatics*, **2016**, 32(21), 3260–3269.

[223] Sciortino G., Garribba E., Rodriguez-Guerra Pedregal J., and Marechal J.-D. *ACS Omega*, **2019**, 4(2), 3726–3731.

[224] Pettersen E. F., Goddard T. D., Huang C. C., Couch G. S., Greenblatt D. M., Meng E. C., and Ferrin T. E. *J. Comput. Chem.*, **2004**, 25(13), 1605–1612.

[225] Alí-Torres J., Maréchal J.-D., Rodríguez-Santiago L., and Sodupe M. *J. Am. Chem. Soc.*, **2011**, 133(38), 15008–15014.

[226] Alonso-Cotchico L., Sciortino G., Vidossich P., Rodriguez-Guerra Pedregal J., Drienovska I., Roelfes G., Lledos A., and Marechal J.-D. *ACS Catal.*, **2019**, 9(5), 4616–4626.

[227] Miki Y., Pogni R., Acebes S., Lucas F., Fernandez-Fueyo E., Baratto M. C., Fernandez M. I., De Los Rios V., Ruiz-Dueñas F. J., Sinicropi A., and others . *Biochem. J.*, **2013**, 452(3), 575–584.

[228] Shi R., Li W., Liu G., and Tang Y. *Chinese J. Chem.*, **2013**, 31(9), 1219–1227.

[229] Piazzetta P., Marino T., and Russo N. *Phys. Chem. Chem. Phys.*, **2014**, 16(31), 16671–16676.

[230] Drienovská I., Alonso-Cotchico L., Vidossich P., Lledós A., Maréchal J.-D., and Roelfes G. *Chem. Sci.*, **2017**, 8(10), 7228–7235.

[231] Hestericová M., Heinisch T., Alonso-Cotchico L., Maréchal J.-D., Vidossich P., and Ward T. R. *Angew. Chem.*, **2018**, 130(7), 1881–1886.

[232] Villarino L., Splan K. E., Reddem E., Alonso-Cotchico L., Souza C., Gutiérrez de, Lledós A., Maréchal J.-D., Thunnissen A.-M. W., and Roelfes G. *Angew. Chem. Int. Ed.*, **2018**, 57(26), 7785–7789.

[233] Mujika J. I., Pedregal J. R.-G., Lopez X., Ugalde J. M., Rodríguez-Santiago L., Sodupe M., and Maréchal J.-D. *Chem. Sci.*, **2017**, 8(7), 5041–5049.

Chapter 2

Objectives

The prediction of the interaction of coordination compounds with proteins is of fundamental importance in many fields like bioinorganic chemistry, biology, pharmacy, medicine and enzyme design. From an experimental perspective, X-ray and NMR methods can provide an accurate description of the system although their application is far from trivial. Other spectroscopic methods, such as EPR, CD and UV-Vis can give insight on the protein region where the metallic species is bound or on the type of amino acid involved in the interaction. However, these instrumental approaches are not able to provide with a complete 3 dimensional description of the systems. Either to predict possible metal mediated binding or enrich partial experimental data, computational methods can represent a valuable approach.

This Ph.D. has been focused on two main objectives. The first one was the validation of an efficient computational strategy to support metallodrug and metalloenzyme design. The central problem was the prediction of the binding site of coordination compounds (a metallodrug itself or a catalytic cofactor) to its biomolecular target. The second one was the integration of instrumental and computational techniques in a hybrid framework to guide for metal binding prediction and model generation.

The specific objectives were: i) the generation and validation of efficient approaches for predicting the binding of coordination compounds with proteins; ii) the integration of spectroscopic and spectrometric data with the *in silico* results and iii) the application of these methodologies to study the speciation and transport of oxidovanadium potential drugs as well as to improve and design artificial metalloenzymes.

Chapter 3

Theoretical Background

The aim of this chapter is to give a general overview of the computational methods and spectroscopic techniques used along his Ph.D. thesis. For each of them a brief theoretical introduction will be given focusing on first principles. Without the presumption to do an exhaustive description of the methods, that can be found in excellent books and reviews, the intention is to provide a comprehensive view of the their key concepts and approximations useful to understand their potentiality and limitations. Purely technical details for the specific calculations and experimental measurements carried out in this thesis will be fully given in the chapters dedicated to the investigation results.

3.1 Quantum Mechanics

3.1.1 Principles

The first principle on which Quantum Mechanics (QM) is based is the postulate stating that the state of a given system can be completely described by a *wavefunction* $\Psi(\mathbf{r}_1, \mathbf{r}_2, ...\mathbf{r}_N, t)$ dependent on the spatial position of the particles $(\mathbf{r}_1, \mathbf{r}_2, ...\mathbf{r}_N)$ and the time (t). The wavefunction contains all the information regarding the specific state observable. In its simplest form, the Ψ can be time-independent and contain all of the information of a given stationary state. A specific Ψ for a microscopic particle is obtained solving the Schrödinger equation, a partial differential equation that was obtained combining the classical description of a wave with the Planck (1900) and Einstein considerations (1905) and the de Broglie wave-particle duality (1924). [1, 2]

Solving the classical wave relation 3.1 using the separation variables with

$u(x,t) = \Psi(x)cos\omega t$, considering the angular velocity $\omega = 2\pi f$ and the phase velocity of a wave, $v = \lambda f$, the equation 3.2 is obtained:

$$\frac{\partial^2 u}{\partial x^2} = \frac{1}{v^2}\frac{\partial^2 u}{\partial t^2} \tag{3.1}$$

$$\frac{\partial^2 \Psi(x)}{\partial x^2} + \frac{4\pi}{\lambda^2}\Psi(x) = 0 \tag{3.2}$$

Taking advantage of the Plank-Einstein ($E = h\nu$) and the de Broglie ($\lambda = {}^{h}/_{mv}$) relations, using the general expression of the total energy of a system [$E_{tot} = {}^{p^2}/_{2m} + V(x)$], λ can be written as $\lambda = {}^{h}/_{\{2m[E-V(x)]\}^2}$ and finally the Schrödinger equation 3.3 can be obtained:

$$-\frac{\hbar^2}{2m}\frac{\partial^2 \Psi(x)}{\partial x^2} + V(x)\Psi(x) = E\Psi(x) \tag{3.3}$$

The equation 3.3 can be formulated as a general eigenvalue problem resulting:

$$\hat{H}\Psi(\mathbf{r}) = E\Psi(\mathbf{r}) \tag{3.4}$$

Following the same method, considering a traveling wave, the time-dependent Schrödinger equation 3.5 can be derived:

$$\hat{H}\Psi(\mathbf{r},t) = i\hbar\frac{\partial \Psi(\mathbf{r},t)}{\partial t} \tag{3.5}$$

The Schrödinger equation itself, in the Copenhagen interpretation, represents a probability amplitude, while its modulus $\mid \Psi \mid^2$ is related to the probability that a particle can be found in a specific region of the space at some instant of time. For this reason an acceptable wavefunction-solution should be square integrable. Solving the equation for Ψ can be used to predict how the particles will behave under the influence of an external potential and with each other. Unfortunately, the Schrödinger equation can be exactly solved for limited cases, in particular for mono-electronic systems. While considering complex potentials or many particles interacting each other, only approximated solutions are possible.

The initial simplification of many bodies problem, used in the time-independent framework, is the Born-Oppenheimer approximation for which the nuclei, having a much higher mass respect to the electrons can be considered as static entities approximating the Schrödinger equation as only function of the electrons. The

Hamiltonian results simplified neglecting the kinetic energy of the nuclei and considering the internuclear distances as a constant parameter (R) in the specific calculation: [1, 2]

$$H\Psi(\mathbf{r};\mathbf{R}) = E\mathbf{R}\Psi(\mathbf{r};\mathbf{R}) \tag{3.6}$$

The parameter could be varied in further calculations determining the potential energy surface (PES) as a function of the nuclear positions. This is a key concept in Quantum Mechanics based studies of reactivity.

Another important simplification used to deal with many electrons systems is the orbital approximation, for which the $\Psi(\mathbf{r_1}, \mathbf{r_2}, ...\mathbf{r_N})$ is considered as a linear combination of independent wavefunctions describing the electrons $\Psi(\mathbf{r}_1, \mathbf{r}_2, ...\mathbf{r}_N) = \Psi(\mathbf{r}_1)\Psi(\mathbf{r}_2)...\Psi(\mathbf{r}_N)$. The single wavefunctions are approximated as hydrogen-like functions (orbitals) influenced for a effective nuclear charge (Z_{eff}) determined by the presence of the other electrons. The atomic orbitals used in this expansion constitute the basis set for the calculation.

In a many body electrons problem, the solution is further complicated by the additional inter-electronic potential energy.

3.1.2 Approximation Methods

As previously stated, the Schrödinger equation can not be solved exactly for many electron systems due to the electron-electron mutual interactions that prevent a solution of the n interacting electrons as a sum of n singly electrons wavefunctions. Thus, approximation methods must be used to solve the Schrödinger equation to almost any desired accuracy. In the following paragraphs the most widely used approximations, the variational principle and the perturbation theory will be briefly presented.

Perturbation Theory

The perturbation theory is built on the consideration that an approximate Hamiltonian $H^{(0)}$ for a given system differs form the real one H by a contribution $H^{(1)}$ defined as *perturbation*. Its application is possible starting from a known simple Hamiltonian and generate most adequate approximate solution for more complicated systems. Considering a generic system the perturbed Hamiltonian, with λ representing a dimensionless parameter that range continuously from 0 (no perturbation) to 1 (full perturbation), results:

$$H = H^{(0)} + \lambda H^{(1)} \tag{3.7}$$

Expanding the H and Ψ as a Taylor series it possible to write the Schrödinger equation as follows:

$$\begin{aligned}&(H^{(0)} + \lambda H^{(1)})(\Psi_0(\mathbf{r}) + \lambda\Psi_1(\mathbf{r}) + \lambda^2\Psi_2(\mathbf{r}) + ...) = \\ &= (E_0(\mathbf{r}) + \lambda\ E_1(\mathbf{r}) + \lambda^2 E_2(\mathbf{r}) + ...)(\Psi_0(r) + \lambda\Psi_1(\mathbf{r}) + \lambda^2\Psi_2(\mathbf{r}) + ...)\end{aligned} \tag{3.8}$$

Equation 3.8 can be solved for such order of λ, λ^L. The general problem could be written as:

$$(\hat{H}^{(0)} - E_0)\Psi_L + (\hat{H}^{(1)} - E_1)\Psi_{L-1} - \sum_{K=2}^{L} E_K\Psi_{L-K} = 0 \tag{3.9}$$

The scalar product of equation 3.9 with Ψ_{L-1} gives the correction λ^L to the unperturbed energy: $E_L = \langle\Psi_0|\hat{H}|\Psi_{L-1}\rangle$. [1, 2]

Variational Principle

Considering Ψ_0 and E_0 representing the exact wavefunction solution and the ground state energy of a given system, the variational principle ensures that substituting Ψ_0 with an alternative trial set of orthonormal wavefunctions, ϕ, $E_\phi \geqslant E_0$. Using a parametrized and combined trial wavefunction, $\phi = \sum_{j=1}^{N} c_j\varphi_j(\alpha, \beta, \gamma)$, the total energy will be dependent on these parameters and the best ground state for ϕ can be obtained by fine tuning the c_j coefficients and α, β and γ parameters. [1, 2]

$$E_\phi - E_0 = \frac{\langle\phi|\hat{H}|\phi\rangle}{\langle\phi|\phi\rangle} - \frac{\langle\psi_0|\hat{H}|\psi_0\rangle}{\langle\psi_0|\psi_0\rangle} = \frac{\langle\phi|\hat{H} - E_0|\phi\rangle}{\langle\phi|\phi\rangle} \geqslant 0 \tag{3.10}$$

3.1.3 Hartree-Fock Method

The first Hartee equation and the subsequent Fock modification accounting for spin symmetry, are based on the orbital approximation and, as many *ab initio* methods, on the iterative Self Consistent Field approach (SCF) to find the best approximate wavefunction for a given system. As commented above, the crucial problem in a many electron system is the electron-electron interaction potential represented in the second sum of the Hamiltanian $H^0 = \sum_{i=1}^{n} h_i + \frac{1}{2}\sum_{i,j}{}' \frac{e^2}{4\pi\varepsilon_0 r_{i,j}}$, where h_i is a hydrogen-like Hamiltonian for the electron i in the field of a nucleus with of charge Z_{eff}. The factor of $1/2$ is to prevent double interactions counting and the prime on the summation excludes terms for which $i = j$ as

electrons do not interact with themselves. The n-electron wavefunction can be separated in orbitals $\Psi_a^0(\mathbf{r}_i;\mathbf{R})$ parametrically dependents on the nuclei position and representing a solution of $h_i\Psi_a^0(\mathbf{r}_i;\mathbf{R}) = E_a^0\Psi_a^0(\mathbf{r}_i;\mathbf{R})$. The Pauli's principle and the spin requirements are satisfied introducing *spinorbitals* obtained as a product of $\Psi_a^0(\mathbf{r}_i;\mathbf{R})$ and a spin function, $\Phi_a(\mathbf{x}_1;R)$ where $\mathbf{x}_i$ contain the spatial and spin state informations. The complete wavefunction is then written as the Slater determinant: [1, 2]

$$\Psi^0(\mathbf{x};\mathbf{R}) = \frac{1}{\sqrt{n!}}\begin{vmatrix} \phi_a(1) & \phi_b(1) & \cdots & \phi_z(1) \\ \phi_a(2) & \phi_b(2) & \cdots & \phi_z(2) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_a(n) & \phi_b(n) & \cdots & \phi_z(n) \end{vmatrix} \tag{3.11}$$

In the Hartree-Fock method (HF) the spin orbitals, giving the best n-electron wavefunction, are built applying the variational principle minimizing the electronic energy ϵ for a specific nuclear configuration. The single orbital HF equation are obtained applying the Fock operator:

$$\begin{aligned} &f_i\phi_a(i) = \varepsilon_a\phi_a(i) \\ &\text{where } f_i = h_i + \sum_u J_u(i) - K_u(i) \end{aligned} \tag{3.12}$$

The Coulomb J_u and the exchange K_u operators take into account the Coulombic repulsion between electrons and the spin correlation, respectively:

$$J_u(i)\phi_a(i) = \left\{ \int \phi_u^*(i+1)\left(\frac{e^2}{4\pi\phi_0 r_{i,j}}\right)\phi_u(i+1)d\mathbf{x}_{i+1} \right\}\phi_a(i) \tag{3.13}$$

$$K_u(i)\phi_a(i) = \left\{ \int \phi_u^*(i+1)\left(\frac{e^2}{4\pi\phi_0 r_{i,j}}\right)\phi_u(i+1)d\mathbf{x}_{i+1} \right\}\phi_u(i) \tag{3.14}$$

The HF equations are computed by the iterative SCF scheme: i) a trial set of spinorbitals are generated and used to formulate the Fock operator; ii) the HF equations are solved obtaining a new set of spinorbitals used to construct a new Fock operator, and so on until achieve a previous defined criterion of convergence.

The actual way to proceed is based on the linear expansion of the unknown ϕ_i over a standard basis function set. In this way the incognits result the simple numerical coefficients reducing the problem on a algebric system of equations.
The *basis-set* choice is of fundamental importance in terms of good balance of accuracy and computational time. For a general molecular system a minimum basis-set describing at least one function is used to represent each occupied orbital of the atoms constituting the system.
The most used functions to describe each orbital are: [1, 2]

Slater-type orbitals (STO) with the general following form:

$$u_{\alpha,n,l,m_l}(r,\theta,\phi) = \frac{1}{\sqrt{(2n)!}}(2\alpha)^{n+1/2}r^{n-1}e^{-\alpha r}Y_{lm_l}(\theta,\phi) \tag{3.15}$$

where α is an empirical parameter and is not necessarily equal to $^{Z}/_{n}$ as in the hydrogen-like orbitals. For polyatomic species, STOs are centered on each of the atoms. For Hartree-Fock SCF calculations the evaluation of the three and four centric integrals results computationally infeasible. Moreover, for different values of n and same values of l and m_l the STOs are not orthogonal to one another. Another deficiency of STOs is that ns-orbitals with $n > 1$ have zero amplitude at the nucleus.

Gaussian-type orbitals (GTO) with the general following form:

$$g_{\alpha,n,l,m_l}(r,\theta,\phi) = Nr^{n-1}e^{-\alpha r^2}Y_{lm_l}(\theta,\phi) \tag{3.16}$$

where N is a normalization constant.

The main advantage of GTOs is that the product of two gaussians at different centers is equivalent to a single Gaussian function centered at a point between the two centers reducing the three- and four-centers integration to a simple two centers problem analytically evaluable. However, the GTOs give a poor representation at small values of r (at the atomic nuclei) and a large basis must be used to achieve an accuracy comparable to STOs. A general procedure, to overcome this limitation is the linear combination of n *primitive Gaussian functions* to fit a Slater type orbital in the so called STO-nG basis-set:

$$\Psi_i = \sum_{k=1}^{M} Ck_i\phi_k \tag{3.17}$$

with ϕ_k represent the *primitive gaussian function*, M the number of functions per orbital and Ck_i are numerical parameters.

The main problem of using a minimal basis-set is due to fixed values of the orbital exponents α of the ϕ_ks, giving orbitals too rigids and unable to contract or expand in different molecular environments. Using a linear combination of orbitals of the same type (retaining the symmetry) but different exponents ξ, the problem can be overcome.

$$\Psi_{nlm_l} = \phi_{nlm_l}(r,\xi_1) + d\phi_{nlm_l}(r,\xi_2) \tag{3.18}$$

where d is a parameter to be determined in the HF procedure.

Using this approach the size of the resulting atomic orbital Ψ_{nlm_l} can range between that specified by $\phi_{nlm_l}(r, \xi_1)$ and $\phi_{nlm_l}(r, \xi_2)$ by varying the d factor. In general, only the valence electrons are described with double-ξ basis-set in the so called split-valence $N - MPG$ basis-set, with N number of gaussian functions used to describe the core electrons and M and P number of gaussian functions (G) used for describe the more compact and the more extended orbitals $\phi_{nlm_l}(r, \xi_i)$, respectively (*e.g.* 6-311g). To the minimal basis-set, orbitals of a higher quantum number of angular-momentum l, can be added accounting for polarization effects of neighbors atoms e.g in the case of hydrogen bonds (*e.g.* 6-311g(d,p)). In addition, further diffuse function (small value of ξ) can complete the basis-set (*e.g.* 6-311g+(d,p)); the use of diffuse functions is particularly useful dealing with electrons far from the nucleus, *e.g.* as lone-pair electrons or electrons in anions.

Finally for heavy elements, in particular for metals, the core electrons could not be explicitly treated as an effective core potential function (ECP or pseudo potential) to reduce basis-set size. A general ECP have the general functional form: [3]

$$U_{ECP}(r) = \sum_i a_i r^{n_i} e^{-\alpha_i r^2} \tag{3.19}$$

where the parameters a_i, n_i and α are obtained reproducing the valence HF energy level and the orbital shape.

In addition, scalar-relativistic effects may be included in effective potentials. [4]

3.1.4 Density functional Theory

Density Functional (DFT), in contrast with HF methods, considers the system energy as a function of the electronic density $\rho_{el}(r)$ avoiding the exact resolution of the Schrödinger equation. [5] The $\rho_{el}(r)$ function is correlated to the electronic wavefunction Ψ_{el} as follows: [6]

$$\rho_{el} = \rho_{el}(r) = \int |\Psi_{el}(r, r_2, \cdots r_N)|^2 dr dr_2 \cdots dr_N \tag{3.20}$$

The relation is guaranteed by the Hohemberg-Kohn theorems: [7]

1. The system energy can be expressed as a functional of the electronic density

$$E = E[\rho_{el}] \quad \text{where} \quad \rho_{el} = \rho_{el}(r) \tag{3.21}$$

2. Defined a given trial density, $\tilde{\rho}_{el}$, in analogy with the variational theorem, the obtained energy will be in any case higher or equal to the exact energy

$$E^o = E[\tilde{\rho}_{el}] \geq E \tag{3.22}$$

The $E[\rho_{el}]$ functional, can be further decomposed applying the Born-Oppenheimer approximation and results:

$$E[\rho_{el}] = T[\rho_{el}]_{el} + V[\rho_{el}]_{nu-el} - V[\rho_{el}]_{el-el} \tag{3.23}$$

The nucleus-electron potential energy ($V[\rho_{el}]_{nu-el}$) can be exactly computed through the Coulomb relation:

$$V[\rho_{el}]_{el-el} = \sum_k \int \frac{Z_k}{|r - r_k|} \rho_{el} dr \tag{3.24}$$

The electron kinetic energy $T[\rho_{el}]_{el}$, depending also on the mutual electronic interactions, presents a higher complexity and can be exactly solved only on the free electrons approximation:

$$T_{ni} = \sum_i \int \frac{1}{2} \bigtriangledown^2 \rho_i dr \tag{3.25}$$

The repulsion electron-electron potential energy ($-V[\rho_{el}]_{el-el}$), in addition to the Coulomb potential, must take into account the non-classic *correlation* and *exchange* effects and for this reason can not be exactly solved:

$$V[\rho_{el}]_{el-el} = V[\rho_{el}]_{coulomb} + V[\rho_{el}]_{exc} + V[\rho_{el}]_{corr} \tag{3.26}$$

The central problem of the DFT approach results in building effective expressions for the termini $T[\rho_{el}]$, $V[\rho_{el}]_{corr}$ e $V[\rho_{el}]_{exc}$. [8] The wide used DFT methods are based on the decomposition of the electronic energy functional 3.23 in a series of universal functionals, for which the exact form is known, and in additional ones specific of the method which include correlation and exchange functionals that also integrate the inter-electronic contribution to the kinetic energy ($\Delta T_{el}[\rho_{el}]$):

$$E[\rho_{el}] = \{T_{ni}[\rho_{el}] + V[\rho_{el}]_{nu-el} - V[\rho_{el}]_{coulomb}\} + \{V[\rho_{el}]_{corr} + V[\rho_{el}]_{exc}\} \tag{3.27}$$

The common functional approximation methods are: *LDA, GGA, MGGA* and *hybrid functionals.*

LDA approximation (*local density approximation*): the V_{corr} and V_{exc} values in a well defined point r^k are only dependent of the the specific ρ_{el}^k value in that point:

$$V_{corr} = V_{corr}[\rho_{el}^k] \qquad V_{exc} = V_{exc}[\rho_{el}^k] \tag{3.28}$$

The well know Slater functional (**S**), bases the V_{exc} calculation on this principle:

$$S = V_{exc} = -\frac{9}{4}\alpha\Big(\frac{3}{4\pi}\Big)^{1/3}\int(\rho_{el}^{\alpha})^{4/3} + (\rho_{el}^{\beta})^{4/3}dr \tag{3.29}$$

where α is an empirical parameter specifically optimized for the functional accuracy. The presence of empirical parameters is a characteristic common to most of the DFT functionals.

GGA approximation (*generalized gradient approximation*): V_{corr} and V_{exc} as well as dependent on ρ_{el}^k, are also dependent on the electronic density gradient:

$$V_{corr} = V_{corr}[\rho_{el}^k, \tfrac{\partial\rho_{el}}{\partial r}|_{r=r^k}] \qquad V_{exc} = V_{exc}[\rho_{el}^k, \tfrac{\partial\rho_{el}}{\partial r}|_{r=r^k}] \tag{3.30}$$

Some example of GGA functionals considering the V_{exc} are the Becke's (B) and the Perdew's and Wang's (PW) functionals, while for V_{corr} the P86, PW91 and LYP. The method is defined by the combination of these functionals of exchange and correlation *e.g.* BLYP, BPW91, *etc...*

MGGA Approximation (*meta-generalized gradient approximation*): represents an extension of the GGA approximation introducing a further dependency of V_{corr} and V_{exc} by the second derivative of the electronic density:

$$\begin{aligned} V_{corr} &= V_{corr}[\rho_{el}^k, \tfrac{\partial\rho_{el}}{\partial r}|_{r=r^k}, \tfrac{\partial^2\rho_{el}}{\partial r}|_{r=r^k}] \\ V_{exc} &= V_{exc}[\rho_{el}^k, \tfrac{\partial\rho_{el}}{\partial r^2}|_{r=r^k}, \tfrac{\partial^2\rho_{el}}{\partial r^2}|_{r=r^k}] \end{aligned} \tag{3.31}$$

Hybrid Functionals: Compute the exchange and correlation contributions using in a weighted manner throughout empirical parameters combining the expressions derived by LDA, GGA and MGGA with HF methods. The most common hybrid functionals, among others, are B3LYP, B3PW91, mPW91, and PBE1PBE (PBE0). The parameters that define the contribute of each terminus are calibrated in an empirical scheme evaluating the functional performance against a reference

dataset. As an example, in the B3LYP functional the energy term of exchange V_{exc}^{B3LYP} is constituted by the 20% V_{exc}^{HF} (exchange Hartree-Fock), by the 8% $V_{exc}^{\mathbf{S}}$ (exchange Slater LDA) and by 72% V_{exc}^{B} (Becke GGA exchange). The correlation part, V_{exc}^{B3LYP} results 19% V_{exc}^{WVN} (WVN LDA correlation) and the 81% V_{exc}^{LYP} (LYP GGA correlation).

The DFT electronic energy, $E(\rho_{el})$, minimization process is based on a self consistence field process (SCF) obtained generating, as a function of the method under consideration, mono-electronic operators, named Kohn-Sham operators ($\hat{K}^i$) dependent on the electronic density:

$$\hat{K}^i(\rho_{el})\Psi_i = \varepsilon_i\Psi_i \qquad \text{where } \varepsilon_i = E_{orbitalic} \tag{3.32}$$

In analogy with the *ab-initio* methods, such as HF, the SCF process allows to optimize the Ψ_i, obtained as a linear sum of basis functions ($\Psi_i = a_k\Phi_k$), tuning the a_k parameters up to reach the self consistence. It can be noted that, in most of the cases, the basis functions used to build the Ψ_i are the same of the *ab-initio* methods.

3.1.5 Solvent Effects

For systems in solution, the medium can be in principle treated describing explicitly the molecules of the solvent or in a implicit way considering the solvent as a continuum model. The addition of explicit solvent molecules sensibly increases the computational cost of the simulation and for this reason the explicit treatment results useful for small systems. Moreover, the number of molecules to be added during the simulation to obtain a good description of the solvent effect still an open question. The continuum model results a good balance between accuracy and computational time. Among these models, one of the most used is the Self-Consistent Reaction Field (SCRF) [9, 10] that treats the solute-solvent interaction as a classical electrostatic problem. The solvent is described as a polarizable continuum medium with a given dielectric constant ϵ and the solute is placed in a proper cavity defined using the vdW radii of the atoms. The charge distribution of the solute induces a polarization at the cavity interface that acts back on it wavefunction producing an electrostatic stabilization in an iteratively process until the mutual polarization achieves the self-consistency. The complete Hamiltonian of the system results the sum of the solute Hamiltonian *in vacuo* $\hat{H}^{solute}$ and a solute-solvent interaction perturbation V^{int} dependent on a solvent polarization response function Q:

$$\hat{H}_{tot} = \hat{H}_{solute} + \hat{V}_{int}(Q) \tag{3.33}$$

The free energy of solvation is computed as a sum of electrostatic, dispersion and repulsion Solute-Solvent interaction and a cavitation term accounting for the energy required to form the cavity in the continuum medium:

$$\Delta G_{sol} = \Delta G_{elec} + \Delta G_{disp} + \Delta G_{rep} + \Delta G_{cav} \tag{3.34}$$

3.1.6 Dispersion Effects

Dispersion interactions, or London forces, are defined as instantaneous dipoles an higher order multipole interactions dominating at regions with scarce electron density probability. At DFT theory level, dispersion is a long-range correlation phenomena difficult to account for. From a practical point of view different empirical model have been proposed to overcome its explicit evaluation such as Grimme's Dn correction to dispersion through the addition of the $^{C_6}/_{R^6} - {}^{C_8}/_{R^8} - {}^{C_{10}}/_{R^{10}}...$ series of terms to the nuclear energy expression. [11, 12] The dispersion empirically derived coefficients are summed over interatomic distances, R_{ij}, modulated by a damping function, $f_{damp}(R_{ij})$, that gradually activates the correction (at a rate characterized by α_6) over a vdW distance, $R_{0,ij}$, the scaling S_6 term being optimized to each Exchange functional: [11, 12]

$$\begin{aligned} E_{disp}^{D2} &= -S_6 \sum_{i,j>i}^{N} f_{damp}(R_{ij}) \frac{C_{6ij}}{R_{ij}^6} \\ f_{damp}(R_{ij}) &= \frac{1}{1 + e^{\alpha_6 (R_{ij}/R_{0,ij} - 1)}} \end{aligned} \tag{3.35}$$

$$\begin{aligned} E_{disp}^{D3} &= -\sum_{n=6,8} S_n \sum_{i,j>i}^{N} \frac{C_n^{ij}}{(R_{ij})^n} f_{damp}(R_{ij}) \\ f_{damp}(R_{ij}) &= \frac{1}{1 + 6(R_{ij}/S_{r,n} R_{0,ij})^{-\alpha_n}} \end{aligned} \tag{3.36}$$

$S_{r,6}$ and S_8 are the custom non-unity parameters fitted for individual functionals. Dispersion effects can also be included in the functional as in the case of the Minnesota functionals that have been parameterized to account for dispersion.

3.2 Molecular Mechanics

Molecular Mechanics (MM) defines the system under analysis basing on the laws of classical physics. For a given molecular system, the potential energy

is calculated as a function of the nuclear positions using a series of empirical parameters defined in a general force fields. The common parameters are available for well-defined functions for bonded and non-bonded interactions. The technique is fast and useful to predict equilibrium geometries, relative energy between different conformers or molecules. [13] Molecular mechanics can be used to compute the potential energy for large molecules or large ensembles of different conformation such as in Molecular Dynamics simulations. However, the technique presents some limitations in terms of bond-breaking and bond-making processes and dealing with metals is rather complicated by the scarce amount of parameters in the available force fields. [14]

MM computes the total energy of the system as a sum of the specific energy contributions of bond stretching, angle bending and dihedral torsion plus the potential energy terms due to non-bonded interactions such as van der Waals (vdW), electrostatics, hydrogen bonds:

$$E_{tot} = E_{str} + E_{bend} + E_{tors} + E_{Hbond} + E_{vdW} + E_{elec} + E_{cross} \tag{3.37}$$

The bonded interaction are generally represented in the most common force fields (such as UFF, MM2, MM3, AMBER, CHARMM, OPLS etc...) [15–22] as Taylor series expansion about the equilibrium position, and specifically described using harmonic functions:

$$\begin{aligned} E(r_{ij}) = \sum_{bonds} k_{r,ij}\left(r_{ij} - r_{e,ij}\right) + \sum_{angles} k_{\theta,ijh}\left(\theta_{ij} - \theta_{ij,eq}\right) \\ + \sum_{torsions}\sum_{n} \frac{1}{2} V_{n,ijhk}\left[1 + cos(n\omega - \lambda)\right] \end{aligned} \tag{3.38}$$

where $r_{ij,eq}$ and $\theta_{ij,eq}$ are the bond and angle equilibrium values, V, n and λ the rotation barrier, the periodicity and the phase factor, respectively.

A valuable and more accurate description of the stretching phenomenon is given by the morse-potential (eq. 3.39):

$$E_{str}(r - r_{eq}) = D\left[1 - e^{k/2D(r - r_{eq})}\right]^2 \tag{3.39}$$

where D is the dissociation energy and k the force constant.
In addition cross terms, accounting for some interactions affecting others, *e.g.* a bent that implies a bond stretching or stretch-stretch, bend-bend, strestch-torsion, *etc...* could be implemented in the force field.

Non-bonded interactions, like vdW and *Hbond*, are generally modeled using Lennard-Jones ($E(r_{ij})^{LJ}$), hard-sphere ($E(r)^{HS}(r)$), square-well ($E(r)^{SW}(r)$) or soft-sphere ($E(r)^{SS}(r)$) potentials:

$$E(r_{ij})^{LJ} = \epsilon_{ij}\left[\left(\frac{r_{e,ij}}{r_{ij}}\right)^{12} - 2\left(\frac{r_{e,ij}}{r_{ij}}\right)^{6}\right] \tag{3.40}$$

$$E(r)^{HS}(r) = \begin{cases}\infty & (r < \sigma)\\ 0 & (\sigma \leqslant r)\end{cases} \qquad E(r)^{SW}(r) = \begin{cases}\infty & (r < \sigma_1)\\ -\epsilon & (\sigma_1 \leqslant r < \sigma_2)\\ 0 & (\sigma_2 \leqslant r)\end{cases} \tag{3.41}$$

$$E(r)^{SS}(r) = \epsilon\left(\frac{\sigma}{r}\right)^{2} \tag{3.42}$$

Electrostatic interactions are generally treated with the classical Coulomb potential:

$$E(r_{ij}) = \sum_{i,j\neq i}\frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \tag{3.43}$$

To obtain the atomic charges that are not physical observables, a general procedure such as the Restrained Electrostatic Potential (RESP) consist in fitting the q_i to reproduce the QM electrostatic potential computed at a large number of grid point around the molecule:

$$\Phi_{ESP}(\mathbf{r}) = \sum_{i}^{nuc}\frac{Z_i}{|\mathbf{R}_i - \mathbf{r}|} - \int\frac{\Psi^*(\mathbf{r}')\Psi(\mathbf{r}')}{|\mathbf{r}' - \mathbf{r}|}d\mathbf{r}' \tag{3.44}$$

This approach has the virtue that such charges reproduce the quantum mechanically determined multipole moments and also the intermolecular interactions with surrounding molecules and is the common procedure used for AMBER force field.

An alternative procedure is represented by the polarizable force fields such as the *Fluctuating Charge Model*.

It must be highlighted that the accuracy and the general applicability of the technique are limited by the number and type of functions that the specific force field accounts for. Moreover, MM requires a series of empirical parameters

(*e.g.* k_{ij}, $r_{e,ij}$, $k_{\theta,ijh}$, $\theta_{e,ijh}$, V_{ijhk}) and, despite their high potentiality and relative low computational cost, most of the force fields lack on appropriate parameterization for specific systems like metal centers, implying a series of limitations in bioinorganic application (see section 3.4).

3.2.1 Metals in ForceFields

The unavailable bonded parameters for metals could be, in principle, derived for the specific system with several methods and tools based on the Seminario method such as FUERZA, [23] MCPB.py [24] or VFFDT. [25] The Seminario method allows to parametrize harmonic bond and angle force field parameters from the QM Hessian matrix of the molecule. The approach consists in computing the reaction force, $\delta\mathbf{F}_i$, due to a small displacement of an atom j respect to the atom i $\delta\mathbf{r}$ in a N atom system:

$$\delta\mathbf{F}_i = -[\mathbf{K}_{ij}]\delta\mathbf{r}_j \tag{3.45}$$

where $[\mathbf{K}]$ is the $3N \times 3N$ Hessian matrix of the molecule.

The three eigenvalues, λ_n^{ij} of $[\mathbf{K}_{ij}]$ are the force constants along the three eigenvectors, ν_n^{ij}. The bond force constant k_{ij} is obtained projecting the three eigenvectors on the bond vector, $\hat{u}^{ij}$:

$$k_{Pi} = \sum_{n=1}^{3} \lambda_n^{ij} |\hat{u}^{ij} \cdot \hat{\nu}_n^{ij}| \tag{3.46}$$

Throughout similar diagonalization procedures of the Hessian matrix, the bend $k_{\theta,ijh}$ force constants could be obtained:

$$\frac{1}{k_{\theta,ijh}} = \frac{1}{R_{ij}^2 k_{Pi}} + \frac{1}{R_{jh}^2 k_{Ph}} \tag{3.47}$$

where R_{ij} and R_{jh} are the lenght of the two bonds forming the ijh angle, and k_{Pi}, k_{Ph} the force constants obtained by the eigenvectors projection onto the perpendicular of the two bonds ij and jk.

3.2.2 Molecular Docking

Protein-ligand docking methods were designed for fast and accurate prediction of the binding sites and interaction modes of small organic molecules to proteins.

Despite their high performances in terms of computational time and wide system exploration, the method considers the macromolecule as a rigid entity neglecting potential conformational changes in solution or induced by binding. [26, 27]

The binding energy calculation is based on simplified MM force fields (Scoring Functions, SF) that take into account non-bonded interactions between the *ligand* (in docking terminology the small molecule) and the protein host. The docking problem is limited to find, generally with a Genetic (GA) or a Montecarlo (MC) algorithm, a series of protein-ligand relative positions (poses) evaluate the energy of each pose throughout the Scoring Function and order them in score ranking.

Scoring Functions

A SF is generally built as a sum of terms representing different interaction type: hydrogen bond (Hbond), electrostatic (elec) or van der Waals (vdW). Such terms could be multiplied by an empirical parameter to fine tune its relative weight. Additionally terms may also be incorporated such as desolvation (desolv), inter-atomic contacts (clashes), ligand torsions (tors), intermolecular ligand interactions or relative to restraints defined (i.e. covalent bonds, coordination interactions etc..). The scoring functions implemented on the available docking software can be grouped in three main categories: force-field-based, empirical, and knowledge-based. Several high quality reviews about Scoring Functions and their performances have been published in the recent years. [28, 29]

Force-Field-Based scoring functions are developed on the top of MM force-fields and, as discussed in the previous paragraph, the different interactions are modeled as atom pair potentials between the protein and the ligand atoms using LJ and Coulombic functional forms. The principal limitation of these kind of SF is that entropic effects, such as the change in energy due to desolvation and solvation of both protein and ligand molecules, are neglected. The *GoldScore* scoring function, described in the following section, is an example of this approach.

Concerning the *Empirical* SF, the fitness of protein-ligand binding is obtained by summing a series of contributions representing the energetic factors in protein-ligand binding. The energetic terms are empirically derived from experimental protein-ligand binding affinities. The wide used *ChemScore* [30] is an example of these kind of SF:

$$\begin{aligned} ChemScore = \nu_0 G_0 + \nu_1 G_{hbond} + \nu_2 G_{metal} + \nu_3 G_{lipo} + \nu_4 G_{rot} + \\ + P_{clash} + c_{internal} P_{internal} \end{aligned} \tag{3.48}$$

where the ν terms are regression coefficients, G represent the various types of contributions to binding and P are penalties which evaluate close contacts and poor internal conformations. Covalent and constraint scores may also be included ($c_{covalent}P_{covalent}$, $P_{constraint}$).

Finally the so called *knowledge-based* scoring functions are built on the *Potential Mean Force* (PME) framework. The Protein-ligand affinity (A^{P-L}) is derived as a sum of distant-dependent intermolecular atomic pair potentials ($\omega_{ij}(r)$):

$$\begin{aligned} A^{P-L} &= \sum_{i}^{Lig.} \sum_{j}^{Prot.} \omega_{ij}(r) \\ \omega_{ij}(r) &= -K_B T ln\left[g_{ij}(r)\right] = -K_B T ln\left[\frac{\rho_{ij}}{\rho^*_{ij}}\right] \end{aligned} \tag{3.49}$$

where the numerical values ρ_{ij}/ρ^*_{ij} reflect the statistical probability for a given pairwise contact. The potentials are derived using a reference set of well characterized protein-ligand complexes e.g. from the Protein Data Bank (PDB). [31] The limitation of the approach is the amount of data available to describe a determined interaction. As an example, the Astex Statistical Potentials (ASP) scoring function is built following this approach.

GoldScore Function

GoldScore was the first SF implemented implemented in GOLD software and, as reported above, is a *force field-based* scoring function. [32, 33] The energy (or Fitness, F) of a determined pose (set of relative protein-ligand positions) is evaluated as a sum of of the intra and intermolecular pairwise potentials relative to vdW, Hbond, and an additional term S_{tors} accounting for the energy change relative to the intra-molecular conformational changes:

$$Fitness = \alpha \cdot S_{hb\_ext} + \beta \cdot S_{vdW\_ext} + \gamma \cdot S_{hb\_int} + \delta \cdot (S_{vdW\_int} + S_{tors}) \tag{3.50}$$

α, β, γ and δ are empirical parameters weighting the relative contribution of each term. By default their values are 1.375 for α, while 1.000 for β, γ and δ. If any constraints and/or covalently bound ligands have been specified, the additional scoring contributions S_{con} and S_{cov} are added to eq. 3.50.

As a first approximation, the terms related to the interaction between two atoms has a *Lennard-Jones 8-4* potential form:

$$S_{i,j} = \sum_{i,j} \left(\frac{A_{ij}}{r^8_{ij}} - \frac{B_{ij}}{r^4_{ij}} \right) \tag{3.51}$$

If the interatomic distance is equal to the sum of vdW radii of the atom pairs the potential utilized is a *Lennard-Jones 12-6*:

$$E_{i,j} = \sum_{i,j}\Big(\frac{C}{r_{ij}^{12}} - \frac{D}{r_{ij}^{6}}\Big) = \sum_{i,j}\Big(\frac{A}{r_{ij}^{12}} - \frac{3}{2}\cdot 0.23 \cdot \frac{I_i I_j \alpha_i \alpha_j}{(I_i + I_j) r_{ij}^{6}}\Big) \tag{3.52}$$

where I_x and α_x are, respectively, the ionization energy and the polarizability of the i and j atoms. The terms A, B, C and D are empirical coefficients present in Gold database. The term S_{hb} is multiplied by a further function ($wt = distance\_wt + angle\_wt$) that takes into account the bond distance and bond angle deviations from an ideal interaction, its terms having the form of a block function:

$$x_{wt}(x, x_{deal}, x_{max}) = \begin{cases} 1, & \text{if } x \le x_{ideal} \\ 1.0 - \dfrac{x - x_{ideal}}{x_{max} - x_{ideal}}, & \text{if } x_{ideal} \le x \le x_{max} \\ 0, & \text{if } x > x_{max} \end{cases} \tag{3.53}$$

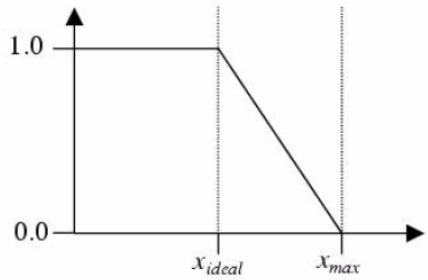

Figure 3.1: Representation of the block function implemented in GOLD

The terms S_{hb_int} and S_{vdW_int}, describing the intra-molecular interactions, are built as the sum of the pairwise intra-molecular energy obtained for each conformation:

$$E_{ij} = \sum_{i,j}\Big(\frac{C}{r_{ij}^{12}} - \frac{D}{r_{ij}^{6}}\Big) \tag{3.54}$$

S_{tors} gives the energy related to a given dihedral ω about the i, j, k, l atoms where n is the periodicity and V the rotational barrier:

$$\mathrm{E}_{ijkl} = \frac{1}{2} V_{ijkl} \sum_{i,j,k,l}\Big[1 + \frac{n_{ijkl}}{\mid n_{ijkl} \mid} cos(\mid n_{ijkl} \mid \cdot \omega_{ijkl})\Big] \tag{3.55}$$

Search Algorithms

A fundamental aspect in docking calculations is the generation of potential docking solutions (relative protein-ligand positions) that can be performed with several search algorithms. Among them, the most used are random walk, Monte Carlo (MC), Simulated Annealing (SA) and Evolutionary Algorithms (EA). The simplest MC approach consists in a *random walk* exploration of the relative position space applying small variation of the ligand degree of freedom (translation, rotation, angle and dihedral etc...). The energy of the new configuration is compared with that obtained for the previous one: if results lower, the configuration is accepted, depending on ΔE:

$$P(\Delta E) = e^{-\frac{\Delta E}{K_B T}} \tag{3.56}$$

where K_B is the Boltzmann constant and T the simulation temperature.

The new conformation is accepted or rejected depending if the computed probability is lower or higher than a random generated number. For high temperature, corresponding to the start of the simulation, mostly of the configurations are accepted. The simulation follows with a series of cycles with different temperatures for which several steps are done. The following cycle starts from the lowest energy configuration and lower temperature determined as:

$$T_i = gT_{i-1} \tag{3.57}$$

where T_i is the temperature of the cycle i, g is a constant ranging from 0 and 1.

Genetic Algorithms (GA) are a particular class of EA that mimics the evolutionary process acting on a set of structures called chromosomes. Due to their fast convergence to reliable solutions have been increasingly used in docking simulations. [34] Each chromosome (or individual) is defined by a series of strings (*alleles*) encoding the relative protein and the ligand positions, their conformation and the mapping of the possible interacting groups. The GOLD docking program, used in this Thesis, is implemented with a fixed population *Steady State with no duplicates* GA in which the *Fitness*, defined a a function of mappings of the fitting points, dihedral of ligand rotable bonds, ligand ring geometries (by flipping ring corners) and the dihedral of protein OH and NH_3^+ groups, is modified/optimized along the simulation. [30, 32, 33] Defined an initial random population of N potential solutions (or individuals), the GA applies two independent *crossover* and *mutation* operators replacing the least x fitnesses members of the original generations. Both operators are applied on randomly selected individuals in a

roulette-wheel-selection based on a weights operator that increases the probability for individuals with highest fitness, introducing an evolutionary pressure. The *crossover* operator copies randomly selected strings from the *parents* combining them into the *children* chromosome producing two new individuals. The *mutation* operator applies a stochastic mutation to a randomly selected string of an individual. The GA automatically terminates following the subsequent application of y genetic runs.

3.3 GaudiMM Platform

GaudiMM (for Genetic Algorithms with Unrestricted Descriptors for Intuitive Molecular Modeling) is a modular platform for molecular modeling developed at the Insilichem group. [35] It combines a Multi-Objective Genetic Algorithm (MOGA) with diverse molecular descriptors. Its grounds consist in transforming any molecular descriptor (i.e. those generally used for analysis of data) as a guiding objective for PES explorations. GaudiMM is built on top of a modular NSGA-II multi-objective genetic algorithm, [36] that in contrast with the previous described GA for which the only optimizable variable is the *Fitness*, it can optimize all the needed variables (*genes*) at once, without compromising any of them over the rest. A MOGA starts with the generation of a random set of individuals. This initial population is with several objectives and assigned a fitness value in order to find the best ones, which are then allowed to recombine and mutate. The results of these variations can be better or worse than their *parents*, but only the best of both the *offspring* and the parental generation ($\mu + \lambda$ strategy) will be selected by the algorithm and propagated to the next generation. After a number of iterations, the initial population will evolve and, eventually, answer with a number of reasonable solutions to the problem (Figure 3.2). [35]

Each possible solution of the problem is represented by an *Individual* object. These objects contain a sorted list of genes and a series of helper methods. A gene describes a molecular feature, such as the topology of the molecule itself, a flexibility model (rotamers, torsions, normal modes), or the spatial orientation.

Each objective is a separate module and class, which defines a single method named evaluate that contains the fitness assessment code. As long as this function returns a quantity that can be maximized or minimized, it will be a valid objective. The same approach works equally for energy estimations, structure-focused optimizations and trivial restraints (distances, angles, dihedrals, surface areas, volume...). As a result, any geometric or energetic parameters that could describe a molecular system can be used as objectives to drive the GA exploration.

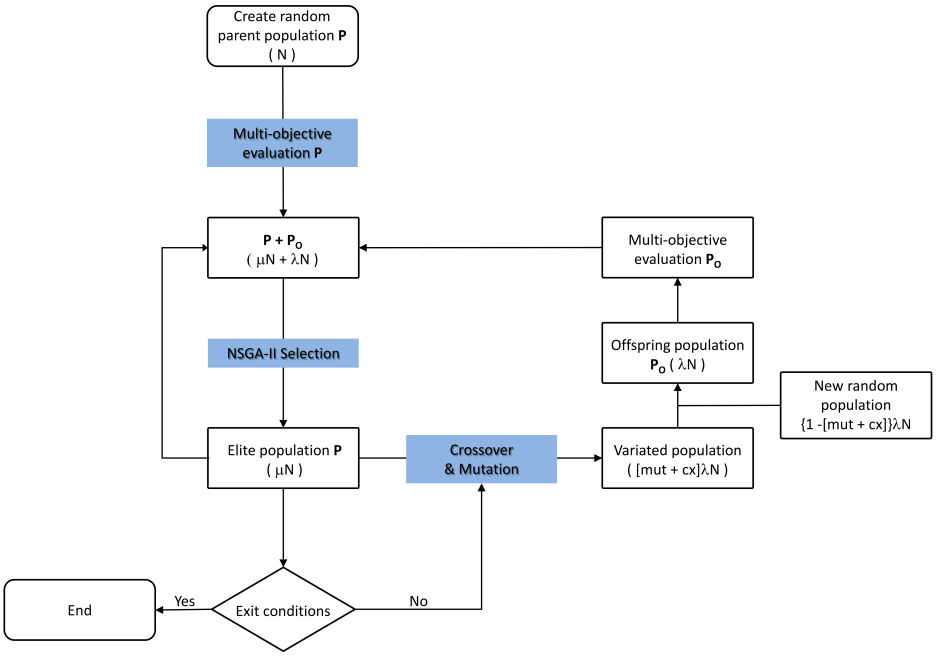

Figure 3.2: Flowchart of the modular NSGA-II multi-objective genetic algorithm implemented in GaudiMM. N is the number of individuals in the initial population P; μ and are respectively related to the number of individuals selected for the next generation and the number of *children* produced at each generation (*offspring* population P_0); *mut* and *cx* are the probabilities associated to *mutation* and *crossover*. Adapted from ref. [35]

3.4 Molecular Dynamics

Molecular Dynamics (MD) simulate the evolution of an interacting particle system along the time. The system and its thermodynamical properties can be treated on one of the *thermodynamics ensembles*: the *microcanonical ensemble* in which an equilibrium system of fixed number of particles, N (or $N_1, N_2, ...N_j$ for a composed system) is considered at constant Energy and Volume (E, N, V); the *canonical ensemble*, where the total energy can change and the Temperature is fixed (T, N, V); the *exothermic-isobaric ensemble*, in which the Pressure is constant (T, N, P); the *grancanonical ensemble* where the chemistry potential μ is fixed (μ, T, V). The most used ensemble in MD simulation is the *microcanonical*

defining an isolated system of interacting particles.

The equation of motion of the single particles is, in the most of the techniques, the classical Newton's equation of motion that can be expressed for N particles considering the position and total momentum vectors $\mathbf{r}(t) = [\mathbf{r}_1(t), \mathbf{r}_2(t), ...\mathbf{r}_N(t)]$ and $\mathbf{p}(t) = \sum_i^N \mathbf{p}_i(t)$ where:

$$\mathbf{p}_i = m_i \frac{d\mathbf{r}_i(t)}{dt} = m_i \mathbf{v}_i \tag{3.58}$$

The total Hamiltonian of the system results:

$$\hat{H} = \hat{T} + \hat{V} = \sum_{i=1}^{N} \frac{\mathbf{p}_i^2}{2m_i} + V(\mathbf{r}) = \frac{1}{2}\mathbf{p} \cdot \mathbf{m}^{-1} \cdot \mathbf{p} + V(\mathbf{r}) \tag{3.59}$$

From eq. 3.59, the thermodynamics properties of the system can be derived if it is known its potential energy:

$$\hat{V} = V[\mathbf{r}_i(t)] \tag{3.60}$$

For the case of *microcanonical ensemble* $V_\mathbf{r}$ can be defined as the the sum of the potential energy of many body interactions:

$$V_\mathbf{r} = \sum_i v_1(\mathbf{r}_i) + \sum_{i<j} v_2(\mathbf{r}_i, \mathbf{r}_j) + \sum_{i<j} v_3(\mathbf{r}_i, \mathbf{r}_j, \mathbf{r}k) + ... \tag{3.61}$$

In the wide used MD techniques, the expression of $V_\mathbf{r}$ is truncated and the potential energy computed only considering two body interactions. Since classical MD is a force field based technique, the different potential used to model intra- and inter-molecular interactions are the same described in the Molecular Mechanics section. In this section, only the methods to solve the the equation of motion will be presented.

3.4.1 Solving the Equations of Motion

The classical Newton's equations of motion for a given isolated system of N particles are defined as: [37]

$$\begin{aligned} m_i \frac{d\mathbf{v}_i(t)}{dt} &= \mathbf{F}_i[\mathbf{r}(t)] = -\nabla V[\mathbf{r}(t)] \\ \frac{d\mathbf{r}_i(t)}{dt} &= \mathbf{v}_i(t) \end{aligned} \tag{3.62}$$

where $\mathbf{F}_i(t)$ is the total force acting on the particle i obtained as the gradient of the potential V respect to its position $\mathbf{r}_i$.

The equation of motion in MD are solved using several numerically algorithms generally based on the finite difference approach that, given the molecular positions, velocities even other dynamic information at time t, the respective quantities at time $t+\delta t$, are computed with a step-by-step process. Among several methods available, the most widely is the the Verlet algorithm obtained as sums or differences of the expansion in Taylor series of the particle coordinates $\mathbf{r}_i(t \pm \delta t)$ about $\mathbf{r}(t)$: [38]

$$\mathbf{r}_i(t+\delta t) = \mathbf{r}_i(t) + \frac{d\mathbf{r}_i(t)}{dt}\delta t + \frac{1}{2}\frac{d^2\mathbf{r}_i(t)}{dt^2}\delta t^2 + \frac{1}{6}\frac{d^3\mathbf{r}_i(t)}{dt^3}\delta t^3 + \mathcal{O}(\delta t^4) \tag{3.63}$$

$$\mathbf{r}_i(t-\delta t) = \mathbf{r}_i(t) - \frac{d\mathbf{r}_i(t)}{dt}\delta t + \frac{1}{2}\frac{d^2\mathbf{r}_i(t)}{dt^2}\delta t^2 - \frac{1}{6}\frac{d^3\mathbf{r}_i(t)}{dt^3}\delta t^3 + \mathcal{O}(\delta t^4) \tag{3.64}$$

Considering the differential expression of the $\mathbf{v}_i$, $\mathbf{a}_i$, and $\mathbf{F}_i$ and neglecting $\mathcal{O}(\delta t^4)$, the series results:

$$\mathbf{r}_i(t+\delta t) = 2\mathbf{r}_i(t) - \mathbf{r}_i(t-\delta t) + \frac{\delta t^2}{m_i}\mathbf{F}_i[\mathbf{r}(t)] \tag{3.65}$$

Following this scheme, the velocities are not needed to compute the trajectories. However, they are fundamental to compute the kinetic energy and thus the total energy. The current velocity can be derived as follow:

$$\mathbf{v}_i(t) = \frac{d\mathbf{r}_i(t)}{dt} = \frac{1}{2\delta t}[\mathbf{r}_i(t+\delta t) - \mathbf{r}_i(t-\delta t)] \tag{3.66}$$

The main disadvantage of the algorithm is the truncation of the Taylor series at the third order as well as the not explicit consideration of the velocities. A modification of the original Verlet is the Leapfrog algorithm that considers integer steps for the coordinates and half steps for velocities: [39]

$$\mathbf{r}_i(t+\delta t) = \mathbf{r}_i(t) + \mathbf{v}_i(t+\frac{\delta}{2}t)\delta t \tag{3.67}$$

$$\mathbf{v}_i(t+\frac{\delta}{2}t) = \mathbf{v}_i(t-\frac{\delta}{2}t) + \frac{d\mathbf{v}_i(t)}{dt}\delta t \tag{3.68}$$

The algorithm stores for each step the current positions, acceleration and the mid step velocities $\mathbf{v}(t - 1/2\delta t)$. The current velocity can be obtained as:

$$\mathbf{v}_i(t) = \frac{1}{2\delta t}[\mathbf{v}_i(t+\frac{1}{2}\delta t) + \mathbf{v}_i(t-\frac{1}{2}\delta t)] \tag{3.69}$$

Similar to the leapfrog, the velocity Verlet computes velocity and position with the same set of variables:

$$\mathbf{r}_i(t+\delta t) = \mathbf{r}_i(t) + \mathbf{v}_i(t)\delta t + \frac{1}{2}\frac{d\mathbf{v}_i(t)}{dt}\delta t^2 \tag{3.70}$$

$$\mathbf{v}_i(t+\delta t) = \mathbf{v}_i(t) + \frac{1}{2}\frac{d\mathbf{v}_i(t)}{dt}\frac{d\mathbf{v}_i(t+\delta t)}{dt}\delta t \tag{3.71}$$

3.4.2 Periodic Boundary Conditions

Dealing with a system in solution such as a solvated molecule in explicit solvent as well as using a continuum model, the system results surrounded by a containing wall. The portion of the medium on the surface will be subject to different forces respect to that in the bulk. The problem can be overcame under Periodic Boundary Conditions (PBC) for which the system is placed on the center of a box replicated on the space to form an infinite lattice. During the simulation, for each molecular motion in the original box corresponds a movement in the same direction of its periodic image on the neighboring boxes. In this way, at the same time that a molecule (*e.g.* a solvent molecule) leaves the central box, other of its images enter through the opposite face conserving the density of the system. The PBC framework avoids to define walls and consequently the surface effects are prevented.
Despite the main advantage of the PBC approach, when using a periodic model for a real macroscopic system the central problem is the correct representation of its thermodynamic properties. In fact if the non-bonded potentials used have longer range than the dimension of the box, the interaction between particles and its own image can induce a deeply increase of the number of interactions to compute. To prevent the problem, a wide used solution consists in using cut-off distances for such intermolecular interactions limiting the separation for $r_{ij} \geqslant r_{cut}$. To reduce the thermodynamic information lost using direct truncation several methodologies to gradually quench the potential have been proposed, such as shifted-potentials: [40–42]

$$v^{SF}(r_{ij}) = \begin{cases} v(r_{ij}) - v_{cut} - \left(\frac{dv(r_{ij})}{dr_{ij}}\right)_{r_{ij}=r_{cut}}, & \text{if } r_{ij} \leq r_c \\ 0, & \text{if } r_{ij} > r_c \end{cases} \tag{3.72}$$

Contextually to cut-offs, long-range corrections can be introduced to partially recover the thermodynamic information.
Analogous problems derive from the long-range Coulomb interactions for such problem the *Ewald Sum* method is used such as in the *Particle Mesh Ewald* (PME) method. [43, 44]

3.5 Hybrid Methodologies

A valuable alternative to pure QM or pure MM based approaches is represented by the so called hybrid methodologies introduced by Honig and Karlplus on 1971, [45] in which part of the system is treated under QM formalism and the rest at classical MM level (Figure 1.10). [46, 47]
These approaches are useful when, for a determined region, is needed a highly accurate description or for which no force field parameters exist (*e.g.*, metal centers in metalloenzymes). In bioinorganics typically the metal ion, the donors of the first coordination sphere and the neighbor amino acid side chains are restricted at QM theory level while the rest of the protein, and, if present, the solvent molecules, are modeled with a force field approach. The hybrid approximation has applicability in the field of statical calculations (QM/MM), used in this thesis work and described in the following paragraphs, as well as for dynamic calculations (QM/MM/MD) described in refs. [46, 48–50].

3.5.1 QM/MM Simulations

The hybrid QM/MM approach, implemented for the first time by Warshel, Levitt [51] and Singh and Kollman, [52] is based on the concept that a chemical system can be partitioned in one region, for which the fine electronic effects are necessary, and a second one that is treated at classical MM level acting as a perturbation. In the simplest case the forces of the classical perturbation exert mechanical constraints on the geometry of the QM region but their effects can be also electrostatic or polarization.

The central problem of the QM/MM approximation is the treatment of the interface between the two regions. Two main approaches are possible, those based on link atoms, in which additional centers are added only on the QM part, or those based on a boundary region in which the additional centers are described at both levels of theory. Finally the last aspect to take into account for continuum systems such as solutes in explicit solvent are the MM boundary conditions that define the energy due to the geometrical constraint imposed to retain the global structure. The total energy of the system result:

$$E_{tot} = E_{QM} + E_{MM} + E_{QM/MM} + E_{boundary} \text{ (ADDITIVE SCHEME)} \quad (3.73)$$

E_{QM} and E_{MM} are the energies of the QM and MM parts computed as isolated systems. $E_{QM/MM}$ is the interaction energy between the two regions and depends on how the interface is defined. It includes all the coupling terms between the two regions and takes into account the vdW interactions and all the modifications of the QM Hamiltonian that exert modifications on the MM part.

Finally $E_{boundary}$ is the energetic term describing all the constraints defined for the external MM region such as fixed coordinates of selected atoms or periodic boundary conditions.

QM/MM Interface For the non-bonded vdW interactions at the QM/MM interface the same formalism described for the MM methods is used while for electrostatic interactions several models are possible. [53] On the *Mechanical Embedding* framework the electronic effects are neglected and the electrostatic interactions are computed only at the MM level using a classical model to describe the QM electrostatic potential. The *Electronic Embedding* in based on including the charge distribution of the MM region as an additional term in the QM Hamiltonian as one-electron operator:

$$\hat{h}_i^{QM/MM} = \hat{h}_i^{QM} - \sum_J^M \frac{e^2 Q_J}{4\pi\epsilon_0 |\mathbf{r}_i - \mathbf{R}_J|} \tag{3.74}$$

where $\mathbf{r}_i$ and $\mathbf{R}_j$ are the position of the electron i and the MM atom J, $\hat{h}_i^{QM}$ the standard one-electron operator and M the number of MM atoms with partial charge Q_J.

The *Polarizable Embedding* also includes the MM polarization due to the QM electrostatic potential. This implies the use of polarizable MM force fields.

Concerning the bonded interaction, in the QM/MM partition, the valence and geometry of both regions must be retained, although for the MM part the simple parametrization represents a valid solution, for the QM part particularly terminations must be included. On the *Molecular Orbital* general schemes [54] localized orbitals are introduced to complete the QM valence and, depending on the approximation, could be placed at the QM (Localized Self-Consistent Field, LSCF) [55] or MM (Hybrid Orbital Approach, GHO) [56] atom. Despite these kind of approaches result an elegant solution, the most used approximation is the simple *link-atom* scheme which completes the QM valence adding additional atoms defining virtual covalent bonds in the QM region with the aim to approximate the real bond with the MM region. Among the several possibilities to define a link atom or a link functional group, the most employed solution is the hydrogen atom placed along the QM/MM broken bond.

ONIOM Scheme

An alternative to the *additive* scheme to compute the QM/MM total energy presented in 3.73 is the *subtractive* model used on the original IMOMM ntegrated Molecular Orbital (MO) + Molecular Mechanic from Morokuma and Maseras, [57] and its evolution ONIOM *Our own N-layered Integrated molecular Orbital and molecular Mechanics* by Morokuma *et al.* implemented in Gaussian software. [58] In ONIOM, the total energy of the whole *real* system is evaluated as the QM energy of the *model* system $E_{QM,model}$ plus the MM energy of the real system $E_{QM,model}$ subtracting the MM energy of the *model* system $E_{QM,model}$:

$$E_{tot,ONIOM} = E_{QM,model} + E_{MM,real} - E_{MM,model} \tag{3.75}$$

In contrast with the additive scheme, a MM evaluation of the model system is required, due to cancellation of the $E_{MM,model}$, only non-bonded and QM/MM boundary parameters are required for its evaluation. Moreover, the scheme presents the main advantage to removes the coupling Hamiltonian $E_{QM/MM}$ avoiding under- or over-counts the QM/MM interactions. [47]

Concerning the Molecular Mechanics energy evaluation, in the Gaussian package are implemented the only UFF, DREIDING and AMBER force fields. Recently our group presented the GARLEEK interface (*Garleek is Automated Resources for Layered Energies using External Keyword*) [59] which, taking advantage of the *external* feature in the Gaussian software, [60] can call an external program, in our implementation the Tinker software, to perform the MM evaluation of the system leaving the complete control of calculation flow to the Gaussian code. GARLEEK expands the application of ONIOM method, allowing the combination of the Gaussian optimization engine with all the force fields implemented in Tinker. [61]

3.6 EPR spectroscopy

3.6.1 Theory

Electron Paramagnetic Resonance (EPR) is the most important instrumental technique to deal with species with unpaired electrons (organic radicals or transition metal complexes). The electron has an intrinsic spin angular momentum of $h[S(S+1)]^{1/2}$ (spin quantum number $S = 1/2$) with components along a determined direction restricted to the values $M_S\hbar$, with $M_S = 1/2, -1/2$. The spin states $|\alpha\rangle$ and $|\beta\rangle$ of the electrons are described by the eigenfunctions $|S, M_S\rangle$ of the operators $\hat{S}^2$ and $\hat{S}_z$:

$$\begin{aligned}\hat{S}^2 |\alpha\rangle &= \hbar^2 S(S+1) |\alpha\rangle \\ \hat{S}^2 |\beta\rangle &= \hbar^2 S(S+1) |\beta\rangle\end{aligned} \tag{3.76}$$

$$\begin{aligned}\hat{S}_z |\alpha\rangle &= {}^1\!/_2 \hbar |\alpha\rangle \\ \hat{S}_z |\beta\rangle &= {}^1\!/_2 \hbar |\beta\rangle\end{aligned} \tag{3.77}$$

For unpaired electrons, the magnetic moment 3.78 associated with their spin angular momentum gives rise to a *paramagnetic* system.

$$\mu = -g_e \frac{e}{2m_e c} \mathbf{S} = -g_e \beta_e \mathbf{S} \left(\frac{\mathbf{S}_z}{\hbar} \right) \tag{3.78}$$

where $g_e = 2.0023$ is the dimensionless g-factor of the free electron and $\beta_e = \frac{e\hbar}{2m_e c}$ is the *Bohr magneton* ($9.27 \times 10^{-28} J/G$ for the electron).

In absence of any magnetic field, the magnetic moments associated with the electron spins, are randomly oriented, and the two energy levels associated to α and β states are degenerate. The application of an external magnetic field $\mathbf{B_0}$ results in an alignment of the magnetic moments associated with the α and β states with respect to the external field vector. Since the parallel orientation is more stable than the anti-parallel one, there is a splitting of the two levels and their energy can be derived by the spin Hamiltonian operator $\hat{H}^S$:

$$\hat{H}^S = -\mu \mathbf{B_0} = g_e \beta \mathbf{B_0} \hat{S} = g_e \beta \left(B_{0,x} \hat{S}_x + B_{0,y} \hat{S}_y + B_{0,z} \hat{S}_z \right) \tag{3.79}$$

If $\mathbf{B_0}$ is oriented along the z axis ($B_{0,z} = B_0$) the expressions for $\hat{H}^S$ and E are:

$$\hat{H}^S = g_e \beta B_0 \hat{S}_z \tag{3.80}$$

$$E = g_e \beta B_0 M_S \tag{3.81}$$

The splitting between the α ($M_S = {}^1\!/_2$) and β ($M_S = -{}^1\!/_2$) energy states is called *Zeeman effect* and is proportional to the magnitude of B_0 (eq. 3.81), as illustrated in Figure 3.3.

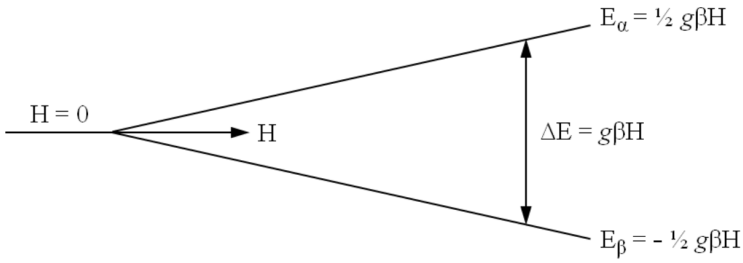

Figure 3.3: Figure of the Zeeman effect in the EPR experiment

The transition between the two levels can be induced by an electromagnetic field of appropriate frequency ν, such as the photon energy $h\nu$ matches ΔE:

$$\Delta E(M_S = 1/2) - E(M_S = -1/2) = g\beta_e B_{0,r} = h\nu \tag{3.82}$$

From equation 3.82, for a given intensity of the magnetic field $B_{0,r}$, the resonance condition depends only on the g factor of the specific system that can diverge from g_e according to the chemical environment.

The EPR concept consists in retaining constant one variable ($B_{0,r}$ or $h\nu$ in equation 3.82) modulating the other, up to induce the spin transition determining the energy gap ΔE and the g factor of the system. Since for experimental reasons results difficult to achieve high sensitivity in microwave modulation, their frequency is usually held constant and the magnetic field is swept linearly. In the most common EPR techniques the microwave value is set around to 9.15 GHz (X band) or 35 GHz (Q band). The spectrum is recorded as first derivative of the absorption as a function of B_0.

3.6.2 EPR spectroscopy of V^{IV} compounds

In most cases, the EPR spectra are complicated by additional magnetic and electric fields exerted by the environment of the electron as in the case of Transition Metals with more than one unpaired electron or in presence of paramagnetic nuclei. The interaction of the unpaired electron with the environment magnetic fields can be described by an additional term to the spin Hamiltonian and leads to further splittings of the energy levels as in the case of vanadium(IV) complexes ($3d^1$ electronic configuration and 99.8% of the ^{51}V isotope). [62, 63]
Considering a small complex in solution at room temperature, their tumbling time is in general smaller than the EPR resolution time ($\sim 50ns$) resulting in a

isotropic orientation distribution of the principal axes respect to the magnetic field vector. The Hamiltonian for a generic $V^{IV}O^{2+}$ complex in these conditions results: [64]

$$\hat{H}^S = g_{iso}\beta B_0 \hat{S}_z + hA_{iso}\hat{S}_z\hat{I}_z \tag{3.83}$$

where g_{iso} is the isotropic g-factor of the complex, and A_{iso} its isotropic hyperfine splitting constant between the spin angular momentum of the electron and that of the vanadium nucleus (^{51}V). $\hat{I}_z$ is the angular nuclear spin momentum operator.

To each eigenvalues M_I of the $\hat{I}_z$ operator it corresponds an eigenstate, $|M_S, M_I\rangle$, with energy: [64]

$$E = g_{iso}\beta B_0 M_S + hA_{iso}M_S M_I \tag{3.84}$$

The term $g_{iso}\beta B_0 M_S$ separates the energy of the electron spin states α and β while the $hA_{iso}M_S M_I$ separates the energy of the components of the nuclear spin angular momentum respect to the z axis (*hyperfine splitting*). Since a nucleus with I spin angular momentum has $(2I+1)$ components along the z axis, with M_I values $-I, -I+1, ..., I-1, I$, the resonance condition (corresponding to the selection rule $\Delta M_S = +1$) is satisfied for $(2I+1)$ values. Natural vanadium contains 99.8% of the ^{51}V isotope with $I = 7/2$; therefore for a $V^{IV}O^{2+}$ species there are $(2I+1) = 8$ values of the external magnetic field which satisfy the resonance condition. In a room temperature spectrum of a $V^{IV}O^{2+}$ complex, 8 transitions with the same intensity are observed, separated by the hyperfine coupling constant A_{iso}. The center of the spectrum corresponds to the g-factor of the examined species (Figure 3.4) and, as discussed above, depends on the chemical environment and on the electronic configuration of the metal ion.

For a solid complex or a species in a frozen solution, characterized by a defined orientation, the g and A values depend on the principal axes orientation and follow the relationships: [64, 65]

$$g_z < g_y \simeq g_x < g_e$$
$$|A_z| < |A_y| \simeq |A_x| > 0$$

This is the case of a rhombic symmetry, that is when the x and y axes are not equivalent. The EPR spectra of a frozen solution will consist of eight

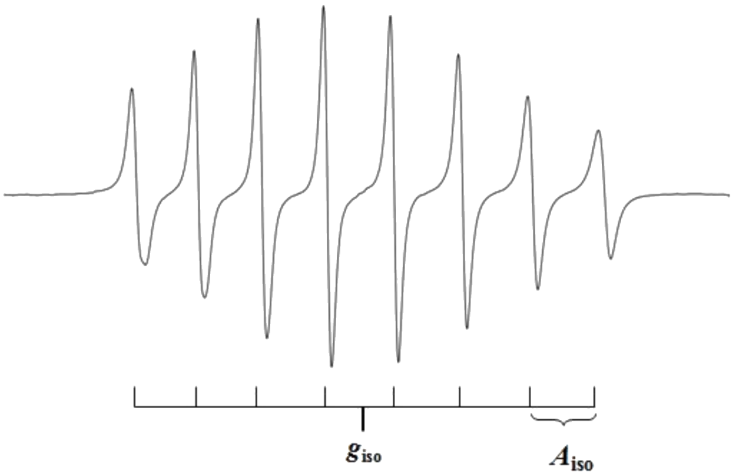

Figure 3.4: Room temperature EPR spectrum of a $V^{IV}O^{2+}$ complex. The positions of the 8 transitions corresponding to the A_{iso} value and the position of g_{iso} are indicated.

equidistant lines separated by A_z and centered around g_z, 8 equidistant lines separated by A_x and centered around g_x, and 8 equidistant lines separated by A_y and centred around g_y (Figure 3.5). For a tetragonal $V^{IV}O^{2+}$ complex, *e.g.* $[VO(H_2O)_5]^{2+}$, x and y axes are coincident and two g and two A values will be measured ($g_{\|} < g < g_e,\ A_{\|} > A_{\perp} > 0$).

Penta-coordinated $V^{IV}O^{2+}$ complexes with bidentate ligands, including those which have ligands with two completely equivalent donor groups like acetylacetonate, can show a significant x, y anisotropy in the EPR spectra recorded in frozen solutions. It was noticed that the rhombicity of the spectrum decreases from penta-coordinated to hexa-coordinated *cis*-octahedral species. A parameter that can be used to define the x, y anisotropy is $|A_x - A_y| = \Delta A_{x,y}$, that follows the order: [66–68] $\Delta A_{x,y}$ (penta-coordinated) $> \Delta A_{x,y}$ (*cis*-octahedral). [67] For a *cis*-octhaedral species the value of $\Delta A_{x,y}$ could be very small when the EPR spectrum is recorded in the X band; in this case, since A_x and A_y are almost equivalent, the resulting spectrum is comparable with a tetragonal system. The x, y anisotropy could be better appreciated in the Q band.

Non-oxido V^{IV} complexes display distinctive spectroscopic features which are discussed in detail in ref. [69].

Concerning the dimeric $(V^{IV}O)_2$ species, the two centers are subject to an

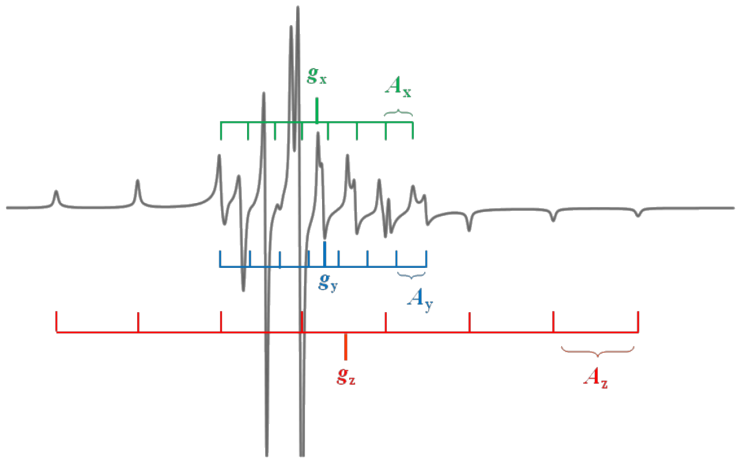

Figure 3.5: Frozen solution EPR spectrum of a $V^{IV}O^{2+}$ complex. The positions of the 8 transitions corresponding to the A_x, A_y and A_z values and the position of g_x, g_y and g_z are indicated

additional mutual magnetic interaction which Hamiltonian can be derived by the linear combination of the spin operator $\hat{S}$ of the two electrons multiplied by the exchange coupling constant J: $\hat{H} = -J\hat{S}_j \cdot \hat{S}_i$. For paramagnetic adjacent centers with $S = 1/2$, the coupling can be diamagnetic for spin state $S = 0$ (singlet) or paramagnetic for $S = 1$ (triplet) with energy difference of $2J$. Thus, if the J value is positive, the triplet state is favored and the resulting interaction is ferromagnetic and an EPR spectrum can be detected. [70] For small energy separations, according to the Boltzmann distribution, the state with higher energy becomes populated for the temperature effect.
The EPR spectra of a triplet system can be interpreted adding the term $\hat{S}D\hat{S}$ to the Hamiltonian presented in eq. 3.80, describing the magnetic moments of the single electron interaction with the surrounding electrons:

$$\hat{H}_S = g_{iso}B_0\hat{S} + \hat{S}D\hat{S} \tag{3.85}$$

For a tetragonal system, even in absence of an external magnetic field, the levels $M_S = \pm 1$ are degenerate and separated from $M_S = 0$ by D (*zero-field splitting*). The degeneracy $M_S = \pm 1$ is removed in an external field and their energy difference is dependent on the intensity of B_0 allowing for the EPR transitions within each pair of levels:

$$\Delta E = D + g_z B_{0,z} = h\nu \tag{3.86}$$

The resulting EPR spectrum is influenced by the D value. In particular for $D > 0$, the spectrum consists of two transitions resonating at values of magnetic fields which differs of $^{2D}/_{g}$. The anisotropic EPR spectrum for the transition $M_S = \pm 1$ consists of four bands centered around g_e, whose separation depends on the molecular interactions (Figure 3.6). The external lines, corresponding to the V=O orientation parallel to the magnetic field direction are separated by $2|D|$, whereas the internal lines corresponding to the perpendicular orientation ($\perp$) are separated by $|D|$, where $|D| = {}^{D}/_{g}$ is the *zero-field splitting* expressed in magnetic field units. For a V^{IV} complex each band is splitted in $(2nI + 1) = 15$ lines, where n is the number of interacting ions, due to the hyperfine coupling of the unpaired electrons with the vanadium nuclei ($n = 2$ for a dinuclear species).

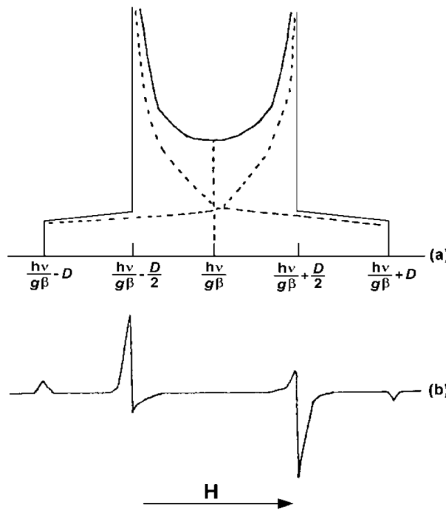

Figure 3.6: (a) Curve of partial (dotted line) and total (solid line) absorption ($\Delta M_S = 1$) for a $S = 1$ system with tetragonal symmetry. (b) First derivative of the total absorption.

Moreover, two sets of 15 lines, partially overlapping, are observed in the forbidden $\Delta M_S = \pm 2$ region, with the magnitude of the coupling constant nearly half of that expected for $A_{\|}$ in the monomeric species with the same equatorial donors.

3.6.3 Prediction of EPR Parameters

The first attempt to predict EPR parameters of $V^{IV}O^{2+}$ compounds is the *additivity relationship* introduced by Wutrich and Chasteen that correlates the hyperfine coupling constant to the number and types of ligands in the equatorial plane of $V^{IV}O^{2+}$. The A_{iso} or A_z values can be estimated as a sum of the empirically derived contribution of each equatorial donor: [71, 72]

$$A_{iso,z} = \sum_{i=1}^{4} A_{iso,z}^{donor,i} \tag{3.87}$$

The contribution to A_{iso} or A_z is approximately inverse to the electron donor capacity of the ligand. Experimental A_z or A_{iso} fall in the range of $\sim 3\times10^{-4}\text{cm}^{-1}$ with respect to the value expected on the basis of the *additivity rule*. [72, 73] In Literature are reported the contributions of donors belonging to the amino, alkoxide, phenolate, aliphatic and aromatic thiolate, hydroxydo Cl^- and SCN^- ions, CO and COO^- groups, imidazole, imine and amide nitrogen. [74–79] However, in most of the cases the empirical relationship is not trivial (for example, when more than one donor set is compatible with the experimental values or with distorted geometries or different ligand orientation respect to the V=O axes where the contribution of the donors to A_z differs from the tabulated values) and more sophisticated methods are required to determine the identity of the analyzed complex.

The EPR parameters can be calculated at DFT level of theory, starting from the optimized structure. Recent advances in computational chemistry have led to the development of several methods based on DFT for the calculation of $\boldsymbol{g}$ and $\boldsymbol{A}$ tensors. [80] In particular, many studies have been devoted to the prediction of hyperfine coupling constants of $V^{IV}O$ complexes. [81–84] The ^{51}V hyperfine coupling tensor $\mathbf{A}$ has three main contributions: the isotropic Fermi contact A^{FC}, the anisotropic or dipolar hyperfine interaction $\mathbf{A}^D$, and one second-order term that arises from spin-orbit (SO) coupling $\mathbf{A}^{SO}$ [69, 85]:

$$\mathbf{A} = A^{FC}\mathbf{1} + \mathbf{A}^D + \mathbf{A}^{SO} \tag{3.88}$$

where $\mathbf{1}$ is the unit tensor, A^{FC} the tensors $\mathbf{A}^D$ and $\mathbf{A}^{SO}$ components $A^D_{\mu\nu}$ and $A^{SO}_{\mu\nu}$ are given by the following equations:

$$A^{FC} = \frac{4\pi}{3} g_e g_N \beta_e \beta_N \langle S_Z \rangle^{-1} \rho_N^{\alpha-\beta} \tag{3.89}$$

$$A^D_{\mu\nu} = \frac{1}{2} g_e g_N \beta_e \beta_N \langle S_Z \rangle^{-1} \sum_{k,l} P^{\alpha-\beta}_{k,l} \langle \Phi_k | \frac{\mathbf{r}^2 \delta_{\mu\nu} - 3\mathbf{r}_\mu \mathbf{r}_\nu}{\mathbf{r}^5} | \Phi_l \rangle \tag{3.90}$$

$$A^{SO}_{\mu\nu} = -\frac{1}{2S} g_e g_N \beta_e \beta_N \sum_{k,l} \frac{\partial P^{\alpha-\beta}_{k,l}}{\partial I_\mu} \langle \Phi_k | h^{SOC}_\nu | \Phi_l \rangle \tag{3.91}$$

g_N, g_e represent the free electron and the nucleus g factors, β_N and β_e are the electron and nucleus magnetons, $\langle S_Z \rangle$ is the expectation value of the electronic spin on the z-axes, $P^{\alpha-\beta}_{k,l}$ is the matrix of nucleus spin density, $\mathbf{r}$ the unpaired electron-nucleus distance, and h^{SOC}_ν represents the spatial part of the one electron SO operator. [85]

The $\mathbf{A}^D$ tensor is always traceless: A^D_x, A^D_y and A^D_z, represent the elements of the diagonalized tensor and their sum being zero:

$$A^D_x + A^D_y + A^D_z = 0 \tag{3.92}$$

The value of the ^{51}V hyperfine constant along x, y and z axes is given by the following equations:

$$A_x = A^{FC} + A^D_x + A^{SO}_x \tag{3.93}$$

$$A_y = A^{FC} + A^D_y + A^{SO}_y \tag{3.94}$$

$$A_z = A^{FC} + A^D_z + A^{SO}_z \tag{3.95}$$

The computed A_z value (A^{calcd}_z) can be compared with the experimental one (A^{exptl}_z) and with those computed by the additivity relationship (A^{expct}_z, eq. 3.87).

The A_{iso} value can be obtained by the following relation:

$$\begin{aligned} A_{iso} = \frac{1}{3}(A_x + A_y + A_z) = A^{FC} + \frac{1}{3}(A^{SO}_x + A^{SO}_y + A^{SO}_z) = \\ = A^{FC} + A^{PC} \end{aligned} \tag{3.96}$$

where $\frac{1}{3}(A_x + A_y + A_z)$ is the pseudo-contact term. [85]

On the ORCA software framework, A_{iso} is obtained in this way. On the other hand, Gaussian software neglects the $A_{\mu\nu}^{SO}$ termini and, therefore, A_{iso} is coincident with A^{FC}:

$$A_{iso} = \frac{1}{3}(A_x + A_y + A_z) = A^{FC} \tag{3.97}$$

Since the contributions $A_z^{D,exptl}$ and $A_z^{SO,exptl}$ can not be isolated from the experimental values, the anisotropic contribution to A along the z axes can be expressed as follows:

$$\begin{aligned} A_{z,aniso} = A_z - A_{iso} = A_z^D + A_z^{SO} - \frac{1}{3}(A_x^{SO} + A_y^{SO} + A_z^{SO}) = \\ = A_z^D + A_z^{SO} - A^{PC} \end{aligned} \tag{3.98}$$

The $A_{z,aniso}^{exptl}$ values can be directly compared with those obtained with the additivity relationship or computed at DFT theory level.

With Gaussian software, that neglects the SO contribution, $A_x^{SO} = A_y^{SO} = A^{SO} = 0$, $A_{z,aniso}$ results:

$$A_{z,aniso} = A_z - A_{iso} = A_z^D \tag{3.99}$$

3.7 Nomenclature

To distinguish the variety of structures considered in the DFT calculations, the IUPAC recommendations for the nomenclature of coordination compounds were taken into account. According to IUPAC, the configuration index of square pyramidal (SPY-5) and octahedral (OC-6) complexes consists of two digits. For SPY-5 the first digit is the priority number of the donor on the C_4 symmetry axis of the idealized pyramid and the second digit is the priority number of the donor trans to that with the lowest priority number in the plane perpendicular to the C_4 axis (if there is more than one of the highest priority ligand in the plane, the priority number of the trans ligand having the largest numerical value is selected). For OC-6 the first digit is the priority number of the ligating atom trans to the ligating atom of priority number 1, and the second digit of the configuration index is the priority number of the ligating atom trans to the most preferred ligand in the plane that is perpendicular to the reference axis. The procedure for assigning priority numbers to the donor atoms is based upon the standard sequence rules developed for chiral carbon compounds by Cahn, Ingold

and Prelog (CIP rules). [86] The presence of polydentate ligands may require the use of primes on some of the numbers in the configuration index. The primes are used to indicate either the donor atoms that are not part of the same polydentate ligand as those that have unprimed priority numbers, or those belong to different parts of a polydentate ligand that are related by symmetry. A primed priority number means that donor atom has lower priority than the same kind of donor atom without a prime on the priority number.

Bibliography

[1] McQuarrie D. A. *Quantum chemistry*, volume 1. University Science Books, 2008.

[2] Atkins P. W. and Friedman R. S. *Molecular quantum mechanics*. Oxford university press, 2011.

[3] Fuentealba P., Preuss H., Stoll H., and Von Szentpály L. A proper account of core-polarization with pseudopotentials: single valence-electron alkali compounds. *Chem. Phys. Letters*, **1982**, 89(5), 418–422.

[4] Dolg M. Relativistic effective core potentials. In *Theoretical and computational chemistry*, volume 11, pages 793–862. Elsevier, 2002.

[5] Szabo A. and Ostlund N. S. *Modern quantum chemistry: introduction to advanced electronic structure theory*. Courier Corporation, 2012.

[6] Hohenberg P. and Kohn W. Inhomogeneous electron gas. *Phys. Rev.*, **1964**, 136(3B), B864.

[7] Kohn W. and Sham L. J. Self-consistent equations including exchange and correlation effects. *Phys. Rev.*, **1965**, 140(4A), A1133.

[8] Perdew J. P. and Schmidt K. Jacob's ladder of density functional approximations for the exchange-correlation energy. In *AIP Conference Proceedings*, volume 577, pages 1–20. AIP, 2001.

[9] Miertus S. and Tomasi J. Approximate evaluations of the electrostatic free energy and internal energy changes in solution processes. *Chem. Phys.*, **1982**, 65(2), 239–245.

[10] Cammi R. and Tomasi J. Remarks on the use of the apparent surface charges (asc) methods in solvation problems: Iterative versus matrix-inversion procedures and the renormalization of the apparent charges. *J. Comput. Chem.*, **1995**, 16(12), 1449–1458.

[11] Grimme S. Semiempirical gga-type density functional constructed with a long-range dispersion correction. *J. Comput. Chem.*, **2006**, 27(15), 1787–1799.

[12] Grimme S., Antony J., Ehrlich S., and Krieg H. A consistent and accurate ab initio parametrization of density functional dispersion correction (dft-d) for the 94 elements h-pu. *J. Chem. Phys.*, **2010**, 132(15), 154104.

[13] Rappe A. K. and Casewit C. J. *Molecular mechanics across chemistry*. University Science Books, 1997.

[14] RappeÂ A. and Casewit C. Molecular mechanics across chemistry university science books, 1997.

[15] Cornell W. D., Cieplak P., Bayly C. I., Gould I. R., Merz K. M., Ferguson D. M., Spellmeyer D. C., Fox T., Caldwell J. W., and Kollman P. A. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.*, **1995**, 117(19), 5179–5197.

[16] Allinger N. L. Conformational analysis. 130. mm2. a hydrocarbon force field utilizing v1 and v2 torsional terms. *J. Am. Chem. Soc.*, **1977**, 99(25), 8127–8134.

[17] Allinger N. L., Yuh Y. H., and Lii J. H. Molecular mechanics. the mm3 force field for hydrocarbons. 1. *J. Am. Chem. Soc.*, **1989**, 111(23), 8551–8566.

[18] Lii J. H. and Allinger N. L. Molecular mechanics. the mm3 force field for hydrocarbons. 2. vibrational frequencies and thermodynamics. *J. Am. Chem. Soc.*, **1989**, 111(23), 8566–8575.

[19] Lii J. H. and Allinger N. L. Molecular mechanics. the mm3 force field for hydrocarbons. 3. the van der waals' potentials and crystal data for aliphatic and aromatic hydrocarbons. *J. Am. Chem. Soc.*, **1989**, 111(23), 8576–8582.

[20] MacKerell Jr A. D., Bashford D., Bellott M., Dunbrack Jr R. L., Evanseck J. D., Field M. J., Fischer S., Gao J., Guo H., Ha S., and others . All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B*, **1998**, 102(18), 3586–3616.

[21] Rappé A. K., Casewit C. J., Colwell K., Goddard III W. A., and Skiff W. M. Uff, a full periodic table force field for molecular mechanics and molecular dynamics simulations. *J. Am. Chem. Soc.*, **1992**, 114(25), 10024–10035.

[22] Jorgensen W. L., Maxwell D. S., and Tirado-Rives J. Development and testing of the opls all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.*, **1996**, 118(45), 11225–11236.

[23] Seminario J. M. Calculation of intramolecular force fields from second-derivative tensors. *Int. J. Quantum Chem.*, **1996**, 60(7), 1271–1277.

[24] Li P. and Merz K. M. Mcpb.py: A python based metal center parameter builder. *J. Chem. Inf. Model.*, **2016**, 56(4), 599–604.

[25] Zheng S., Tang Q., He J., Du S., Xu S., Wang C., Xu Y., and Lin F. Vffdt: A new software for preparing amber force field parameters for metal-containing molecular systems. *J. Chem. Inf. Model.*, **2016**, 56(4), 811–818.

[26] Weill N., Therrien E., Campagna-Slater V., and Moitessier N. Methods for docking small molecules to macromolecules: a user's perspective. 1. the theory. *Curr. Pharm. Des.*, **2014**, 20(20), 3338–3359.

[27] Campagna-Slater V., Therrien E., Weill N., and Moitessier N. Methods for docking small molecules to macromolecules: A user's perspective. 2. applications. *Curr. Pharm. Des.*, **2014**, 20(20), 3360–3372.

[28] Liu J. and Wang R. Classification of current scoring functions. **2015**, 55(3), 475–482.

[29] Huang S.-Y., Grinter S. Z., and Zou X. Scoring functions and their evaluation methods for protein–ligand docking: recent advances and future directions. *Phys. Chem. Chem. Phys.*, **2010**, 12(40), 12899–12908.

[30] Verdonk M. L., Cole J. C., Hartshorn M. J., Murray C. W., and Taylor R. D. Improved protein–ligand docking using gold. *Proteins: Struct., Funct., Bioinf.*, **2003**, 52(4), 609–623.

[31] Berman H. M., Westbrook J., Feng Z., Gilliland G., Bhat T. N., Weissig H., Shindyalov I. N., and Bourne P. E. The protein data bank. *Nucleic Acids Res.*, **2000**, 28(1), 235–242.

[32] Jones G., Willett P., and Glen R. C. Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation. *J. Mol. Biol.*, **1995**, 245(1), 43–53.

[33] Jones G., Willett P., Glen R. C., Leach A. R., and Taylor R. Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.*, **1997**, 267(3), 727–748.

[34] Guan B., Zhang C., and Ning J. Genetic algorithm with a crossover elitist preservation mechanism for protein–ligand docking. *AMB Express*, **2017**, 7(1), 174.

[35] Rodríguez-Guerra Pedregal J., Sciortino G., Guasp J., Municoy M., and Maréchal J.-D. Gaudimm: A modular multi-objective platform for molecular modeling. *J. Comput. Chem.*, **2017**, 38(24), 2118–2126.

[36] Deb K., Pratap A., Agarwal S., and Meyarivan T. A fast and elitist multiobjective genetic algorithm: Nsga-ii. *IEEE transactions on evolutionary computation*, **2002**, 6(2), 182–197.

[37] Allen M. P. and Tildesley D. J. *Computer simulation of liquids*. Oxford university press, 2017.

[38] Verlet L. Computer" experiments" on classical fluids. i. thermodynamical properties of lennard-jones molecules. *Phys. Rev.*, **1967**, 159(1), 98.

[39] Forest E. and Ruth R. D. Fourth-order symplectic integration. *Physica D*, **1990**, 43(1), 105–117.

[40] Brooks III C. L., Pettitt B. M., and Karplus M. Structural and energetic effects of truncating long ranged interactions in ionic and polar fluids. *J. Chem. Phys.*, **1985**, 83(11), 5897–5908.

[41] Bergdorf M., Peter C., and Hünenberger P. H. Influence of cut-off truncation and artificial periodicity of electrostatic interactions in molecular simulations of solvated ions: A continuum electrostatics study. *J. Chem. Phys.*, **2003**, 119(17), 9129–9144.

[42] Steinbach P. J. and Brooks B. R. New spherical-cutoff methods for long-range forces in macromolecular simulation. *J. Comput. Chem.*, **1994**, 15(7), 667–683.

[43] Essmann U., Perera L., Berkowitz M. L., Darden T., Lee H., and Pedersen L. G. A smooth particle mesh ewald method. *J. Chem. Phys.*, **1995**, 103(19), 8577–8593.

[44] Fischer N. M., Maaren P. J., van, Ditz J. C., Yildirim A., and Spoel D., van der. Properties of organic liquids when simulated with long-range lennard-jones interactions. *J. Chem. Theory Comput.*, **2015**, 11(7), 2938–2944.

[45] Honig B. and Karplus M. Implications of torsional potential of retinal isomers for visual excitation. *Nature*, **1971**, 229(5286), 558.

[46] Senn H. M. and Thiel W. Qm/mm methods for biomolecular systems. *Angew. Chem. Int. Ed.*, **2009**, 48(7), 1198–1229.

[47] Chung L. W., Sameera W. M. C., Ramozzi R., Page A. J., Hatanaka M., Petrova G. P., Harris T. V., Li X., Ke Z., Liu F., Li H.-B., Ding L., and Morokuma K. The oniom method and its applications. *Chem. Rev.*, **2015**, 115(12), 5678–5796.

[48] Li P., Roberts B. P., Chakravorty D. K., and Merz J., Kenneth M. Rational design of particle mesh ewald compatible lennard-jones parameters for +2 metal cations in explicit solvent. *J. Chem. Theory Comput.*, **2013**, 9(6), 2733–2748.

[49] Vidossich P., Lledós A., and Ujaque G. First-principles molecular dynamics studies of organometallic complexes and homogeneous catalytic processes. *Acc. Chem. Res.*, **2016**, 49(6), 1271–1278.

[50] Vidossich P. and Magistrato A. Qm/mm molecular dynamics studies of metal binding proteins. *Biomolecules*, **2014**, 4(3), 616.

[51] Warshel A. and Levitt M. Theoretical studies of enzymic reactions: dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *J. Mol. Biol.*, **1976**, 103(2), 227–249.

[52] Singh U. C. and Kollman P. A. A combined ab initio quantum mechanical and molecular mechanical method for carrying out simulations on complex molecular systems: Applications to the ch3cl+ cl- exchange reaction and gas phase protonation of polyethers. *J. Comput. Chem.*, **1986**, 7(6), 718–730.

[53] Reuter N., Dejaegere A., Maigret B., and Karplus M. Frontier bonds in qm/mm methods: A comparison of different approaches. *J. Phys. Chem. A*, **2000**, 104(8), 1720–1735.

[54] Warshel A. and Levitt M. Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *J. Mol. Biol.*, **1976**, 103(2), 227–249.

[55] Ferré N., Assfeld X., and Rivail J.-L. Specific force field parameters determination for the hybrid ab initio qm/mm lscf method. *J. Comput. Chem*, **2002**, 23(6), 610–624.

[56] Gao J., Amara P., Alhambra C., and Field M. J. A generalized hybrid orbital (gho) method for the treatment of boundary atoms in combined qm/mm calculations. *J. Phys. Chem. A*, **1998**, 102(24), 4714–4721.

[57] Maseras F. and Morokuma K. Imomm: A new integrated ab initio+ molecular mechanics geometry optimization scheme of equilibrium structures and transition states. *J. Comput. Chem.*, **1995**, 16(9), 1170–1179.

[58] Svensson M., Humbel S., Froese R. D., Matsubara T., Sieber S., and Morokuma K. Oniom: a multilayered integrated mo+ mm method for geometry optimizations and single point energy predictions. a test for diels- alder reactions and pt (p (t-bu) 3) 2+ h2 oxidative addition. *J. Phys. Chem.*, **1996**, 100(50), 19357–19363.

[59] Pedregal J. R.-G., Funes-Ardoiz I., Sciortino G., Sánchez-Aparicio J.-E., Ujaque G., Lledós A., Maréchal J.-D., and Maseras F. Garleek: Adding an extra flavor to oniom. *J. Comput. Chem.*, **2019**, 40(2), 381–386.

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

[61] Ponder J. W. and others . Tinker: Software tools for molecular design, 2004.

[62] Ballhausen C. J. and Gray H. B. The electronic structure of the vanadyl ion. *Inorg. Chem.*, **1962**, 1(1), 111–122.

[63] Mabbs F. Some aspects of the electron paramagnetic resonance spectroscopy of d-transition metal compounds. *Chem. Soc. Rev.*, **1993**, 22(5), 313–324.

[64] Gerson F. J. a. weil, j. r. bolton j. e. wertz. electron paramagnetic resonance: Elementary theory and applications. wiley-interscience, new york (1994). xxi 568 pages, £66.00/$91.95. isbn 0-471-57234-9. *Magn. Reson. Chem.*, **1995**, 33(1), 80–80.

[65] Drago R. S. Physical methods in chemistry. Technical report, 1977.

[66] Garribba E., Micera G., Panzanelli A., and Sanna D. Electronic structure of oxovanadium (iv) complexes of α-hydroxycarboxylic acids. *Inorg. Chem.*, **2003**, 42(13), 3981–3987.

[67] Lodyga-Chruscinska E., Micera G., and Garribba E. Complex formation in aqueous solution and in the solid state of the potent insulin-enhancing vivo2+ compounds formed by picolinate and quinolinate derivatives. *Inorg. Chem.*, **2011**, 50(3), 883–899.

[68] Micera G. and Garribba E. The effect of trigonal bipyramidal distortion of pentacoordinate vivo2+ species on their structural, electronic and spectroscopic parameters. *Eur. J. Inorg. Chem.*, **2011**, 2011(25), 3768–3780.

[69] Sanna D., Sciortino G., Ugone V., Micera G., and Garribba E. Nonoxido viv complexes: Prediction of the epr spectrum and electronic structure of

simple coordination compounds and amavadin. *Inorg. Chem.*, **2016**, 55(15), 7373–7387.

[70] Garribba E., Micera G., Lodyga-Chruscinska E., and Sanna D. Oxovanadium (iv) complexes with pyrazinecarboxylic acids: The coordinating properties of ligands with the (naromatic, coo–) donor set. *Eur. J. Inorg. Chem.*, **2006**, 2006(13), 2690–2700.

[71] Wüthrich K. Elektronenspinresonanz-untersuchungen von vo2+-komplexverbindungen in wässeriger lösung ii. *Helv. Chim. Acta*, **1965**, 48(5), 1012–1017.

[72] Chasteen N. D. Vanadyl (iv) epr spin probes inorganic and biochemical aspects. In *Biological magnetic resonance*, pages 53–119. Springer, 1981.

[73] Smith II T. S., LoBrutto R., and Pecoraro V. L. Paramagnetic spectroscopy of vanadyl complexes and its applications to biological systems. *Coord. Chem. Rev.*, **2002**, 228(1), 1–18.

[74] Tolis E. J., Teberekidis V. I., Raptopoulou C. P., Terzis A., Sigalas M. P., Deligiannakis Y., and Kabanos T. A. The effect of charged axial ligands on the epr parameters in oxovanadium (iv) compounds: an unusual reduction of the az (51v) values. *Chem. Eur. J.*, **2001**, 7(12), 2698–2710.

[75] Hamstra B. J., Houseman A. L., Colpas G. J., Kampf J. W., LoBrutto R., Frasch W. D., and Pecoraro V. L. Structural and solution characterization of mononuclear vanadium (iv) complexes that help to elucidate the active site structure of the reduced vanadium haloperoxidases. *Inorg. Chem.*, **1997**, 36(21), 4866–4874.

[76] Jakusch T., Buglyó P., Tomaz A. I., Pessoa J. C., and Kiss T. Thiolate-s as anchoring donor in the binary and ternary vo (iv) complexes of mercaptopropionylglycine. *Inorganica Chim. Acta*, **2002**, 339, 119–128.

[77] Smith T. S., Root C. A., Kampf J. W., Rasmussen P. G., and Pecoraro V. L. Reevaluation of the additivity relationship for vanadyl- imidazole complexes: Correlation of the epr hyperfine constant with ring orientation. *J. Am. Chem. Soc.*, **2000**, 122(5), 767–775.

[78] Garribba E., Lodyga-Chruscinska E., Micera G., Panzanelli A., and Sanna D. Binding of oxovanadium (iv) to dipeptides containing histidine and cysteine residues. *Eur. J. Inorg. Chem.*, **2005**, 2005(7), 1369–1382.

[79] Tasiopoulos A. J., Troganis A. N., Evangelou A., Raptopoulou C. P., Terzis A., Deligiannakis Y., and Kabanos T. A. Synthetic analogues for oxovanadium (iv)–glutathione interaction: An epr, synthetic and structural study of oxovanadium (iv) compounds with sulfhydryl-containing pseudopeptides and dipeptides. *Chem. Eur. J.*, **1999**, 5(3), 910–921.

[80] Neese F. Prediction of molecular properties and molecular spectroscopy with density functional theory: From fundamental theory to exchange-coupling. *Coord. Chem. Rev.*, **2009**, 253(5), 526–563.

[81] Munzarová M. L. and Kaupp M. A density functional study of epr parameters for vanadyl complexes containing schiff base ligands. *J. Phys. Chem. B*, **2001**, 105(50), 12644–12652.

[82] Aznar C. P., Deligiannakis Y., Tolis E. J., Kabanos T., Brynda M., and Britt R. D. Ese-endor study and dft calculations on oxovanadium compounds: Effect of axial anionic ligands on the 51v nuclear quadrupolar coupling constant. *J. Phys. Chem. A*, **2004**, 108(19), 4310–4321.

[83] Saladino A. C. and Larsen S. C. Density functional theory calculations of the electron paramagnetic resonance parameters for vo2+ complexes. *J. Phys. Chem. A*, **2003**, 107(11), 1872–1878.

[84] Neese F. Metal and ligand hyperfine couplings in transition metal complexes: The effect of spin–orbit coupling as studied by coupled perturbed kohn–sham theory. *J. Chem. Phys.*, **2003**, 118(9), 3939–3948.

[85] Neese F. O. *ORCA - An Ab Initio, DFT and Semiempirical Program Package, Version 3.0.* Max-Planck-Institute for Chemical Energy Conversion: Mülheim a. d. Ruhr, 2013.

[86] Damhus T., Hartshorn R., and Hutton A. Nomenclature of inorg. chem.: Iupac recommendations 2005. *Chem. Int.*, **2005**.

Chapter 4

Methodological Results

In this chapter, new approaches to accurately predict the interaction with proteins of coordination complexes with bioinorganic relevance (i.e. metallodrugs but also physiologically occurring metal species), using protein-ligand docking are presented. The first part of the work aims at a very general framework valid for most of nowadays protein-ligand docking programs although our proof of concept has been done with the commercially available software GOLD. The second one presents an implementation of the *Insilichem* modeling suite GaudiMM of chemical rules for metal coordination geometry prediction.

4.1 Improving GOLD for Coordination Docking

The main problem in docking involving metalloligand is the absence of clear parameters to account for the metal ions. To extend the docking software GOLD to predict the binding of a metal complex to a protein, it was assumed that, from a pure computational point of view, hydrogen bond functions could be related to coordination bonds. The hard work consists in generating the convenient atom types and scoring functions to adapt this idea to docking software without using any geometrical constraint or energy restraint. To test this approach, the model has been applied to a set of 39 high-quality X-ray structures with transition and main group metal complexes bound to proteins with a unique coordination bond. The results show an excellent agreement between the predicted and the crystallographic structures, indicating that the method could be used as a new tool to predict metal complexes-proteins interactions also when the X-ray structure is not available. [1]

A second step along the way was to generalize the framework for the binding of metalloligands with more than one vacant site. The benchmark was augmented to 64 structures formed by main group- and transition metal-containing ligands with various coordination numbers (4-6), geometries (square-planar, tetrahedral, trigonal bipyramidal, square-pyramidal, and octahedral), and accessible coordination sites. Again the results were particularly promising. We also applied the methodology to five case studies involving the binding of pharmacologically active metal species to proteins: (i) the anticancer copper(II) complex $[Cu^{II}(Br)(2\text{-hydroxy}-1\text{-naphthaldehydebenzoylhydrazine})(indazole)]$ to human serum albumin (HSA); (ii) one of the active species of antidiabetic and antitumor vanadium compounds, $V^{IV}O^{2+}$ ion, to carboxypeptidase; (iii) the antiarthritic species $[Au^{I}(PEt_3)]^{+}$ to HSA; (iv) the antitumor oxaliplatin to ubiquitin; (v) the antitumor ruthenium(II) compound RAPTA-PentaOH to cathepsin B. [2]

4.2 GaudiMM Objective for Metal Ions Binding Site Prediction

In the former part on this line of studies, directionality of the coordination bonds were introduced in dockings. However, this approach is still quite limiting and does not account for specific coordination functions. With the aim to generalize the range of applicability of metal binding site prediction, a new objective of has been implemented on our multipurpose molecular modeling suite, GaudiMM. The benchmark of the approach was performed on the prediction of the bare metal ion binding sites in proteins. Unlike the previous discussed extension for the protein-ligand software GOLD, the new molecular descriptor (or objective) reproduces coordination geometries of metal ions that can bind the protein with any number of coordination vacancies. Moreover, this objective does not rely on structural templates or bond strength parametrization extracted from large protein data sets, but on pure coordination rules (mainly geometrical) that are structure agnostic. The whole process is based on finding poses that satisfy metal-derived geometrical rules without needing sequence or fine electronic inputs (e.g. from quantum mechanical calculations). The approach was benchmarked on 105 X-ray structures and for all of them the correct binding location in the biological scaffold was found. [3]

This work clearly shows how explicitly considering the geometric particularities of the first coordination sphere of the metal in a docking process provides excellent results and could be a step forward in metal related biomedical and biotechnological fields.

Bibliography

[1] Sciortino G., Rodríguez-Guerra Pedregal J., Lledós A., Garribba E., and Maréchal J.-D. Prediction of the Interaction of Metallic Moieties with Proteins: An Update for Protein–Ligand Docking Techniques. *J. Comput. Chem.*, **2018**, 39(1), 42–51.

[2] Sciortino G., Garribba E., and Marechal J.-D. Validation and Applications of Protein–Ligand Docking Approaches Improved for Metalloligands with Multiple Vacant Sites. *Inorg. Chem.*, **2018**, 58(1), 294–306.

[3] Sciortino G., Garribba E., Rodriguez-Guerra Pedregal J., and Marechal J.-D. Simple Coordination Geometry Descriptors Allow to Accurately Predict Metal–Binding Sites in Proteins. *ACS Omega*, **2019**, 4(2), 3726–3731.

Chapter 5

Application of the Methods

In this chapter the applications and the extension of the developed methodology in the field of metallodrug and artificial metalloenzyme design are discussed. The interaction with proteins *in vivo* plays a key role in metallodrug bioactivity, because it can influence their transport in the bloodstream, the cell uptake, and finally their mechanism of action.

Concerning metalloenzyems, the interaction of the cofactor with its host, and the identification of catalytically active protein-cofactor geometries is a crucial aspect.

Therefore, the knowledge of metal interactions with bioligands is fundamental to understand metallodrugs or metalloenzyme mode of action as well as to design novel compounds or new artificial metal-based enzymes.

5.1 Vanadium Metallodrugs

In general, two types of binding are expected in a (metal complex)-protein system. (i) A covalent or cooordination binding, when the protein replaces with one or more side chain residues the organic ligand L or a weak ligand such as water from metal first coordination sphere; for example, species with composition *cis*-VOL_2−(Protein) are formed from *cis*-VOL_2−(H_2O) after the replacement of the equatorial water molecule by a His-N or Asp/Glu-COO donor. (ii) A noncovalent binding, when the complex VOL_2 interacts through secondary interactions, such as van der Waals and hydrogen contacts, with the accessible groups on the protein surface.

Concerning the first case, most theoretical approaches struggle to get a hand

on though those discussed in the previous chapter open new avenues. They have been here applied to the study of the interaction of two potential $V^{IV}OL_2$ drugs with anti-diabetic activity with lysozyme, a wide used model protein, for which EPR spectroscopy provides partial data. The results demonstrate that integrating EPR with blind protein-ligand dockings allows to predict the binding site and also the enantiomeric preference of V^{IV} species. [1]

In a second showcase, a combination of EPR and docking techniques was applied to study the noncovalent interaction between square pyramidal VOL_2 complexes and lysozyme. EPR spectroscopy is a good indicator of the strength of such an interaction because at room temperature a *rigid limit* spectrum is expected when the noncovalent binding is strong enough to block the metal species on the protein surface and a *slow tumbling* or an *isotropic* spectrum when the binding is weak and the complexes are free to rotate in solution. Docking methods, after an accurate parametrization, provide valuable insights into the possible binding sites of V species to the protein surface. Comparing the computational and EPR data was found a good indicator to measure the strength of this interaction allowing to mark the transition from a *rigid limit* (strong interaction) to a *slow tumbling* or an *isotropic* (weak or no interaction). The reliability of the computational results was further evaluated through QM/MM refinement. [2]

The results show that convenient updates of protein-ligand docking programs allow to satisfactorily reproduce X-ray structures of metal complexes that interact with throughout direct coordination bond(s) to proteins and predict with blind procedures relevant low energy modes of binding. They also demonstrate that the combination of the docking methods with spectroscopic data (EPR spectroscopy, in this study) could represent the interactions of metal containing ligands with proteins and could have a general applicability in this field including for paramagnetic species. We believe that this studies fills an important gap in the possibility to develop metallodrugs by incorporating molecular modeling into the toolkit of designers and so expand methodological practice common in the organic drug design community to those fields where transition metals are used as a pillar for developing therapeutic agents.

5.1.1 Interaction with biomolecular carriers

The pharmacological effect of vanadium metallodrugs is highly related to the interaction with bioligands, *i.e.* the transport in bloodstream by human serum transferrin, the reduction of V^V to $V^{IV}O$ species in the cytosol, where they bind to hemoglobin, and the inhibition of phosphatase by $H_2V^VO_4^-$ is the basis of the V antidiabetic activity, while the binding of V^{IV} or V^V species to holo or apo-transferrin could promote their cellular uptake. Unfortunately, the lack of three-dimensional knowledge of these adducts with proteins limits the

development of new active molecules. Thus, the study of these systems appears to be fundamental to increase the knowledge of the type of interaction, the number of the species interacting, and the region of the proteins where the binding occurs.

Dealing with V^{IV} compounds, EPR gives insights into the nature of the equatorial donors of $V^{IV}O^{2+}$ complexes. ESI (electrospray ionization) and MALDI (matrix-assisted laser desorption/ionization) represent the most powerful mass spectrometries (MS) to explore metallodrug-protein interactions. They give information on the number and stoichiometry of the metal-protein adducts formed, and ESI-MS in particular has been used to study the reactivity of several metallodrugs. Since EPR and ESI-MS do not allow identifying the exact region of the protein where the metal is bound, the information provided by these techniques can be completed by computational methods and, in particular, docking calculations.

In the following sections an integrated EPR/ESI-MS/Computational strategy is applied to: i) decipher the interaction of $V^{IV}O^{2+}$ ion and several $V^{IV}O$ potential metallodrugs with myoglobin, [3] and ii) the interaction with ubiquitin and lysozyme of two $V^{IV}O$ quinolone compounds, with different size and net charge evaluating the effect of these variables on the binding capabilities. The stabilization of the secondary hydrogen bonds and van der Waals contacts on the stability of the adducts formed was also evaluated. [4]

The results, providing new information on the binding of potential vanadium drugs, could pave the way for a rational design of new pharmacologically active compounds and highlight a series of features that must be tuned to optimize the interaction strength of V species with the target protein(s).

5.2 Key Factor of a Diel-Alderase αRep protein

The developed computational technique has been also applied in the field of artificial metallenzymes design, in particular to a specific homodimer variant, A3, of the αRep family of artificial repeat proteins. The interest of this scaffold relies on the stability of its interdomain cavity and in the possibility to express the protein as a single chain bidomain metalloenzymes, offering the opportunity to selectively link a co-factor to one of its mutated subunits (A3'). A single mutation ensures the covalent coupling of a 1:1 Cu(II)-phenanthroline or Cu(II)-terpyridine complex as a catalytic center within the interdomain cavity able to accommodate two substrates of the Diels-Alder reaction. Four variants of the artificial Diels-Alderases, have been synthesized coupling both co-factors to two different mutated amino acids, and their activity rationalized by means of our integrated instrumental/computational strategy suggesting the key factors of the selectivity. These results open the way to a rational catalyst optimization and may constitute valuable tools for a more efficient and selective artificial biocatalysis. [5]

5.3 Oxaliplatin Interaction with Insulin

The search for anticancer platinum compounds to replace cisplatin, which has several side effects, has brought to the second generation Pt drugs with lower toxicity. Among others, these compounds include carboplatin and oxaliplatin. Oxaliplatin is a metallodrug especially effective against metastatic colorectal cancer and is much less nephro- and ototoxic than cisplatin. In addition, oxaliplatin can be used to treat cisplatin-resistant tumors. Oxaliplatinum contains the labile dianionic oxalate ligand and in the blood gives $[Pt(dach)]^{+}$, with two free coordination positions, that can be attacked by the serum bioligands, among which proteins. This phenomenon has been pointed out to be responsible of off-target interactions that have major therapeutic consequences including side effects, loss of bioavailability and emergence of resistance.

Insulin is one of the prototypical target Pt-based drugs since it is involved in bioavailability reduction and might also determine resistance in certain cancer lines.

In the following sections the interaction of oxaliplatin with insulin was examined through an integrated multilevel computational strategy to provide structural insights into this system. To do so, the initial structures are predicted by blind protein–metalloligand docking calculations optimized to account for a metal-containing species, and then refined using a Molecular Dynamics (MD) and Quantum Mechanics/Molecular Mechanics (QM/MM) integrated protocol.

The results provide new information like conformational changes occurring upon binding and potential effects on the biological functions of the protein. The proposed multiscale approach for predicting the binding of metallodrugs to protein hosts, its implications in conformational changes of bioligands and the potential effects on their biological functions, open a new avenue towards applying similar strategies to a wide ensemble of metallodrug–protein/peptide systems for which no structural data are available. [6]

Bibliography

[1] Sciortino G., Sanna D., Ugone V., Micera G., Lledos A., Marechal J.-D., and Garribba E. Elucidation of Binding Site and Chiral Specificity of Oxidovanadium Drugs with Lysozyme through Theoretical Calculations. *Inorg. Chem.*, **2017**, 56(21), 12938–12951.

[2] Sciortino G., Sanna D., Ugone V., Lledos A., Marechal J.-D., and Garribba E. Decoding Surface Interaction of $V^{IV}O$ Metallodrug Candidates with Lysozyme. *Inorg. Chem.*, **2018**, 57(8), 4456–4469.

[3] Sciortino G., Sanna D., Ugone V., Maréchal J.-D., and Garribba E. Integrated ESI–MS/EPR/Computational Characterization of the Binding of Metal Species to Proteins: Vanadium Drug–Myoglobin Application. *Inorg. Chem. Front.*, **2019**, 6, 1561–1578.

[4] Sciortino G., Sanna D., Ugone V., Maréchal J.-D., Alemany-Chavarria M., and Garribba E. Effect of Secondary Interactions, Steric Hindrance and Electric Charge on the Interaction of $V^{IV}O$ Species with Proteins. *New J. Chem.*, **2019**. DOI: 10.1039/C9NJ01956A.

[5] Di Meo T., Kariyawasam K., Ghattas W., Valerio-Lepiniec M., Sciortino G., Marechal J.-D., Minard P., Mahy J.-P., Urvoas A., and Ricoux R. Functionalized Artificial Bidomain Proteins Based on an α–Solenoid Protein Repeat Scaffold: A New Class of Artificial Diels–Alderases. *ACS Omega*, **2019**, 4(2), 4437–4447.

[6] Sciortino G., Sánchez-Aparicio J.-E., Pedregal J. R.-G., Garribba E., and Maréchal J.-D. Computational Insight into the Interaction of Oxaliplatin with Insulin. *Metallomics*, **2019**, 11(4), 765–773.

Chapter 6

Conclusions

This Ph.D. aimed at expanding molecular modeling applicabilities in the challenging field of the prediction of the interaction of inorganic and organometallic compounds with proteins. Such objective has been reached throughout a series of computational developments and studies of applicative cases. Altogether, it provides with major proof-of-concept clear evidence on the power of computation in this particular bioinorganic area.

Computational efforts have been primarily given to the expansion of protein-ligand dockings towards the inorganic world. Two different pieces of software were appended with capabilities for dealing with metalloligands.

One is the commercial software GOLD, one of the most cited program in the chemical biology community. After generated new scoring functions optimized for metal ions and metalloligand, the approach was benchmarked over 64 structures formed by main group- and transition metal-containing ligands with various coordination numbers, geometries, and accessible coordination sites. The agreement between computed and experimental structures was astonishingly good. Since the basis of the scoring function is to adapt hydrogen bond terms to simulate coordination bonds, this approach is virtually adaptable to almost all docking programs. [1, 2]

To better account for real coordination rules in docking procedure, a new objective has been implemented in our molecular modeling suite, GaudiMM. This molecular descriptor explicitly accounts for the geometrical rules of the metals in determining where inorganic moieties could interact with proteins. Its accuracy has been benchmarked on the prediction of the bare metal ion binding sites in proteins, blindly reproducing a dataset of 105 X-ray structures, and obtaining for

all the entries the correct binding location in the biological scaffold. [3]
All these software developments clearly demonstrate that convenient updates in docking methodologies can arm them with high predictivenss for metalloligand-protein interactions and could have a general applicability in metal mediated biomedical and biotechnological fields.

In the following part of the study, the capabilities of metal friendly docking methods have been applied to a series of systems. A first focus was given on bridging computation with Electron Paramagnetic Resonance spectroscopy (EPR) [4, 5] and Electro Spray Ionization Mass Spectrometry (ESI-MS): two methods that provide insight on the interaction of metals with proteins but do not provide full three dimensional resolution of the entire system. Indeed, dealing with paramagnetic V^{IV} compounds, EPR gives informations about the nature of the equatorial donors of the complexes, while ESI-MS on the number and stoichiometry of the metal-protein adducts formed. [6] The combined computational-experimental methodology has been applied to the prediction of covalent and non-covalent binding modes of potential V^{IV} antidiabetic drugs with lysozyme, myoglobin, ubiquitin, human serum transferrin and albumin. [6–10] The relevance of the results obtained by this mixture between experimental and theoretical information clearly open the way to a rational design of new vanadium based pharmacologically active compounds but also how to guide a vast number of fields in bioinorganic chemistry.

This is actually what is shown in the work performed on artificial metalloenzymes presented in this work. One key element in the efficient construct of Artificial metalloenzymes is a proper interaction between an coordination compounds (the artificial cofactor) and the protein host. The strategies developed in this Ph.D. were therefore applied on such kind of systems. [11, 12] In particular, it has applied to study a single chain bidomain αRep metalloenzymes, A3-A3', covalent coupled with 1:1 Cu(II)-phenanthroline or Cu(II)-terpyridine complex as a catalytic center with DielsAlderases activity. The catalysis promoted by four different variants has been rationalized by means of our integrated instrumental/computational strategy unveiling the key factors of its selectivity. [13] The results offer the opportunity to rational optimize the catalyst and may constitute valuable information to design metalloenzyme based on the same αRep library with new catalytic activity. [13]

In the final part of the thesis, focus is given in overcoming the limitation of docking simulations when it comes to simulate the dynamical nature of the proteins. When ligand binding has impact on the dynamical behavior of the host, docking experiments tend to be far too limited. In those cases in which the coordination sphere of the metal could eventually adapt during the binding

process, multilevel approaches, combining Molecular Mechanics based strategies such as protein-ligand docking or classical Molecular Dynamics with pure QM and QM/MM approaches, provides with an efficient framework to better understand the binding mechanism. The same strategy can be applied to investigate the response of a protein to the interaction with a metal as demonstrated for the binding of cisplatin with insulin. [14] The results provide with new information like conformational changes occurring upon binding and their potential effects on the biological functions of the protein.

Bibliography

[1] Sciortino G., Rodríguez-Guerra Pedregal J., Lledós A., Garribba E., and Maréchal J.-D. Prediction of the Interaction of Metallic Moieties with Proteins: An Update for Protein–Ligand Docking Techniques. *J. Comput. Chem.*, **2018**, 39(1), 42–51.

[2] Sciortino G., Garribba E., and Marechal J.-D. Validation and Applications of Protein–Ligand Docking Approaches Improved for Metalloligands with Multiple Vacant Sites. *Inorg. Chem.*, **2018**, 58(1), 294–306.

[3] Sciortino G., Garribba E., Rodriguez-Guerra Pedregal J., and Marechal J.-D. Simple Coordination Geometry Descriptors Allow to Accurately Predict Metal–Binding Sites in Proteins. *ACS Omega*, **2019**, 4(2), 3726–3731.

[4] Sciortino G., Sanna D., Ugone V., Micera G., Lledos A., Marechal J.-D., and Garribba E. Elucidation of Binding Site and Chiral Specificity of Oxidovanadium Drugs with Lysozyme through Theoretical Calculations. *Inorg. Chem.*, **2017**, 56(21), 12938–12951.

[5] Sciortino G., Sanna D., Ugone V., Lledos A., Marechal J.-D., and Garribba E. Decoding Surface Interaction of $V^{IV}O$ Metallodrug Candidates with Lysozyme. *Inorg. Chem.*, **2018**, 57(8), 4456–4469.

[6] Sciortino G., Sanna D., Ugone V., Maréchal J.-D., and Garribba E. Integrated ESI–MS/EPR/Computational Characterization of the Binding of Metal Species to Proteins: Vanadium Drug–Myoglobin Application. *Inorg. Chem. Front.*, **2019**, 6, 1561–1578.

[7] Sciortino G., Sanna D., Ugone V., Maréchal J.-D., Alemany-Chavarria M., and Garribba E. Effect of Secondary Interactions, Steric Hindrance and Electric Charge on the Interaction of $V^{IV}O$ Species with Proteins. *New J. Chem.*, **2019**. DOI: 10.1039/C9NJ01956A.

[8] Ugone V., Sanna D., Sciortino G., Maréchal J.-D., and Garribba E. Interaction of Vanadium(IV) Species with Ubiquitin: A Combined Instrumental and Computational Approach. *Inorganic Chemistry*, **2019**, 58(12), 8064–8078.

[9] Sanna D., Ugone V., Sciortino G., Buglyó P., Bihari Z., Parajdi-Losonczi P. L., and Garribba E. $V^{IV}O$ Complexes with Antibacterial Quinolone Ligands and their Interaction with Serum Proteins. *Dalton Trans.*, **2018**, 47(7), 2164–2182.

[10] Banerjee A., Dash S. P., Mohanty M., Sanna D., Sciortino G., Ugone V., Garribba E., Reuter H., Kaminsky W., and Dinda R. Chemistry of Mixed–Ligand Oxidovanadium(IV) Complexes of Aroylhydrazones Incorporating Quinoline Derivatives: Study of Solution Behavior, Theoretical Evaluation and Protein/DNA Interaction. *J. Inorg. Biochem.*, **2019**, 199, 110786.

[11] Rodríguez-Guerra J., Alonso-Cotchico L., Sciortino G., Lledós A., and Maréchal J.-D. Computational Studies of Artificial Metalloenzymes: From Methods and Models to Design and Optimization. *Artificial Metalloenzymes and MetalloDNAzymes in Catalysis: From Design to Applications*, **2018**.

[12] Alonso-Cotchico L., Sciortino G., Vidossich P., Rodriguez-Guerra Pedregal J., Drienovska I., Roelfes G., Lledos A., and Marechal J.-D. Integrated Computational Study of the Cu–Catalyzed Hydration of Alkenes in Water Solvent and into the Context of an Artificial Metallohydratase. *ACS Catal.*, **2019**, 9(5), 4616–4626.

[13] Di Meo T., Kariyawasam K., Ghattas W., Valerio-Lepiniec M., Sciortino G., Marechal J.-D., Minard P., Mahy J.-P., Urvoas A., and Ricoux R. Functionalized Artificial Bidomain Proteins Based on an α–Solenoid Protein Repeat Scaffold: A New Class of Artificial Diels–Alderases. *ACS Omega*, **2019**, 4(2), 4437–4447.

[14] Sciortino G., Sánchez-Aparicio J.-E., Pedregal J. R.-G., Garribba E., and Maréchal J.-D. Computational Insight into the Interaction of Oxaliplatin with Insulin. *Metallomics*, **2019**, 11(4), 765–773.

Acknowledgements

Firstly, I would like to express my sincere gratitude to my directors Prof. Jean-Didier Maréchal and Prof. Eugenio Garribba for the continuous support of my Ph.D study and related research, for their patience, motivation, and knowledge. Their guidance helped me in all the time of research. I could not have imagined having better directors and mentors for my Ph.D study.

Besides my directors, I would like to thank Dr. Daniele Sanna who gave access to the laboratory and research facilities. Without his precious support it would not be possible to conduct this research.

My sincere thanks also goes to Prof. Agustí Lledós and Prof. Gregori Ujaque who provided me a unique opportunity to join their research field, and for their insightful comments and encouragement.

All the collaborators during this Ph.D. period are also acknoledgwed, notably Jaime Rodriguez-Guerra (Charité - Universitätsmedizin Berlin), Jose Emilio Sanchez (Universitat Autònoma de Barcelona), Lur Alonso-Cotchico (Rijksuniversiteit Groningen), Valeria Ugone (Università degli studi di Sassari), the group of Péter Buglyó and Norbert Lihi (Debreceni Egyetem), the group of João Costa Pessoa (Universidade de Lisboa), the group of Xavier Verdaguer and Antoni Riera (Instituto de Investigación Biomédica Barcelona), the group of Jean-Pierre Mahy and Rémy Ricoux (Université Paris-Sud), the group of Eugenio Vázquez (Centro Singular de Investigación en Química Biolóxica e Materiais Moleculares), the group of Rupam Dinda (National Institute of Technology, Rourkela), the group of Feliu Maseras (Institut Català d'Investigació Química), the group of Mercè Capdevila and Oscar Palacios (Universitat Autònoma de Barcelona), the group of Rosa M. Ortuño and Ona Illa (Universitat Autònoma de Barcelona), and the group of Sankar Prasad Rath (Indian Institute of Technology Kanpur).

The Generalitat de Catalunya (2017SGR1323), Spanish MINECO (grant CTQ2017-87889-P), Fondazione di Sardegna (grant FdSGarribba 15) and Regione Autonoma della Sardegna (grant RASSR79857) are kindly acknowledged for the financial contribution, and Universitat Autonoma de Barcelona for support this Ph.D.