

UNIVERSITA' DEGLI STUDI MILANO-BICOCCA

FACOLTA' DI SCIENZE MM.FF.NN.

DOTTORATO IN BIOTECNOLOGIE INDUSTRIALI

XXIV Ciclo



**COMPUTATIONAL MODELING OF THE ENZYMIC
ACTIVITIES OF BIOMOLECULES AT DIFFERENT SCALES:
FROM QUANTUM MECHANICAL REACTION STUDIES TO
SYSTEMIC UNDERSTANDING OF CELL BEHAVIOR**

Coordinatore: Prof. Marco Ercole Vanoni

Tutor: Prof. Piercarlo Fantucci

VALENTINA BARBIERI

MATRICOLA n° 040425

Anno accademico 2010/2011

To my parents,
even though they would deserve
much more than this

“Don't panic”
The hitchhiker's guide to the galaxy – D.Adams

Contents

1. Abstract	8
2. Riassunto	14
3. Introduction	20
3.1 The role of modeling in cell biology	20
3.2 Aim of the thesis	23
3.3 References	24
4. Quantum mechanic studies of reactivity mechanism of bioinorganic molecules	25
4.1 Methods of Quantum Chemistry	26
4.1.1. Density Functional Theory methods	28
4.2. DFT characterization of key intermediates in thiols oxidation catalyzed by amavadin	33
4.2.1 The biological question	32
4.2.2. Computational details	36
4.2.3. Results and discussions	37
4.2.3.1. <i>Geometry and electronic structure of reduced and oxidized forms of amavadin</i>	38
4.2.3.2. <i>DFT optimization of the protonated form of $[V(hidpa)_2]^-$</i>	40
4.2.3.3. <i>Characterization of reaction intermediates (adducts) formed when MMA reacts with $[V(hidpa)_2H]$</i>	43
4.2.3.4. <i>Characterization of reaction intermediates formed when MMA reacts with $[V(hidpa)_2]^-$</i>	47
4.2.3.5. <i>Final mechanistic considerations on thiol oxidation</i>	49
4.2.4. Conclusions	51

4.2.5. References	52
4.3. DFT investigation of the copper coordination to the putative cell binding site of angiogenin	54
4.3.1. The biological question	54
4.3.2 Computational methods	57
4.3.3. Results and Discussion	58
4.3.3.1 <i>Geometries and relative energies of isomers of Ac-PHREN-NH₂/Cu²⁺·H₂O and Ac-PHRQN-NH₂/Cu²⁺·H₂O complexes</i>	58
4.3.4. Conclusions	71
4.3.5. Acknowledgments	75
4.3.6. References	75
5. Bistability and microdomains in cell fate decision	
76	
5.1 The biological question	77
5.2. Theoretical approach	81
5.2.1 Stochasticity	83
5.2.2 Smoldyn software	87
5.2.2.1 <i>Basic theory</i>	87
5.2.2.2 <i>The algorithm</i>	90
5.2.2.3 <i>Molecular diffusion</i>	91
5.2.2.4. <i>Surfaces</i>	92
5.2.2.5. <i>Zeroth-order reactions</i>	95
5.2.2.6. <i>Unimolecular reactions</i>	95
5.2.2.7. <i>Bimolecular reactions</i>	96
5.3 Results	97
5.3.1. Ras-Sos activation signaling three-dimensional model	97
5.3.2. Analogy between fermions and bistable systems	101
5.4. Conclusions	102
5.5 References	103
6. Dynamic model of the G1/S phase transition in the Gq/11	

protein signaling network	104
6.1. The biological problem	104
6.2. Theoretical approach	109
6.2.1. Ordinary differential equations	109
6.2.2. Virtual Cell	111
6.2.2.1. CVODE	112
6.2.2.2. <i>Parameter estimation</i>	114
6.2.2.3. <i>Experimental data</i>	114
6.2.2.4. <i>Methodological issues</i>	115
6.2.2.5. <i>Simulations settings</i>	116
6.3. Results	117
6.3.1. Activation of MAPKs	117
6.3.2. Activation of c-jun and c-fos gene	119
6.3.3. Activation of some components of the G1/S regulating machine	120
6.4 Conclusions and perspectives	121
6.5. References	122
7. Final conclusions	123
8. Acnowledgments	125
9. Publications	128

1. Abstract

Computational modeling of the enzymatic activities of biomolecules at different scales: from quantum mechanical reaction studies to systemic understanding of cell behavior.

The aim of the thesis is the development of computational models of the enzymatic activity of biomolecules at different scales. Parallel investigations have been carried out at a quantum level to study the reactivity of an enzyme from an electronic point of view, and at a systemic level using simulation techniques to determine the role of enzymes in the network of cellular reactions.

Starting from the lowest complexity level, the thesis begins with two computational studies with the aims of describing the molecular mechanism of the catalytic reaction between amavadin and methyl mercaptoacetate and the structural coordination between angiogenin and copper ion, respectively. Since in both cases transition metals are involved, Density Functional Theory (DFT) computations and Quantum Theory of Atoms in Molecules (QTAIM) analysis of the electron density were used.

The first study regards the electronic description of the enzymatic mechanism by which molecule amavadin, an unusual octa-coordinated V^{IV} complex isolated from *Amanita muscaria* mushrooms, catalyzes the oxidation of some thiols to the corresponding disulfides.

An experimental work by G.da Silva et al. proposed an inner-sphere mechanism for the reaction (*J. Am. Chem. Soc.* 1996, **118**:7568–7573.) but the electronic mechanism was not identified.

In the first step of our investigation, the stereoelectronic features of the V^{IV}

(inactive) and V^V (active) states of amavadin were determined by means of DFT. Then, the formation of the V^V complex with methyl mercaptoacetate (MMA), which has been chosen as a prototypical substrate, has been characterized both thermodynamically and kinetically. DFT results reveal that protonation of V^V amavadin at a carboxylate oxygen not directly involved in the V coordination, favors MMA binding into the first coordination sphere of vanadium, by substitution of the amavadin carboxylate oxygen with that of the substrate and formation of an $S-H\cdots O$ hydrogen bond interaction. The latter interaction can promote SH deprotonation and binding of the thiolate group to vanadium. The kinetic and thermodynamic feasibility of the V^V -MMA intermediates formation is in agreement, along with electrochemical experimental data, also with the biological role exerted by amavadin. Finally, the presence of an ester functional group as an essential requisite for MMA oxidation has been rationalized. Moreover, the results proposed an indirect evidence on the role of the vanadium (in its d^0 active state) based catalyst protonation.

The second studies focused on the structural and spectroscopic properties of the complexes formed by Cu^{2+} and the peptide fragments Ac-PHREN-NH₂, which encompasses the putative cell binding domain of angiogenin, as well as its Ac-PHREN-NH₂ variant. Analysis of structures, relative energies and EPR parameters has allowed to conclude that the metal coordination environment at pH 8 is formed by a nitrogen atom of His, two deprotonated amide groups, a water molecule and an oxygen atom from the COO⁻ side chain of Glu, in nice agreement with recent experimental results [*Dalton Trans*, 2010, **39**:10678]. Moreover, DFT results allowed to reveal that the Glu sidechain of the Ac-PHREN-NH₂ peptide is coordinated in equatorial position, fully disclosing the effects of Cu^{2+} binding on the structural

properties of this key angiogenin portion.

We have also investigated the configuration space of the E→Q mutated system Ac-PHRQN-NH₂/Cu²⁺·H₂O. In this case, computational results led to the conclusion that the H₂O molecule is coordinated in equatorial position and the oxygen atom of the carbonyl group of glutamine is weakly coordinated in apical position.

Using as a reference the recent experimental results reported by Bonomo and collaborators [*Dalton Trans*, 2010, **39**:10678], we have used classical MM/MD calculations followed by DFT optimizations to explore the configurational space of the Ac-PHREN-NH₂/Cu²⁺·H₂O complex.

Both of these DFT studies have been carried out in the laboratory of professor Piercarlo Fantucci, Department of Biotechnology and Biosciences, University of Milan-Bicocca.

These two works lead to the following publications:

“DFT characterization of key intermediates in thiols oxidation catalyzed by amavadin”, *Dalton Transaction*, 40 (30): 7704-7712, 2011.

“Copper coordination to the putative cell binding site of angiogenin”. A DFT investigation., *Inorg.Chemistry. Accepted*.

Acquired some modeling skills at the molecular level, we decided to add another layer of complexity to our investigation and we decided to test if a correlation could exist between the peculiar structure of a protein and a biological effect at a cellular level.

The fundamental biological system of the Ras/Sos signaling activation pathway has been chosen for our study.

Based on two different previous studies we developed a model that can

correlate bistability and the microdomains clustering in this small network. In fact Das et al. (*Cell* 2009, **136**:337-351, 2009) reported that the Sos-dependent Ras activation in lymphocytes, beyond stimulus, causes a digital (*on/off*) response that implies a positive feedback loop in the regulation mechanism of SOS activity. The deterministic model developed by the authors demonstrates that for low or high levels of Sos there is only one possible state correlated to low or high levels of active Ras, respectively. For intermediate values of SOS, three states of Ras activity are generated (two states are stable and could be simultaneously observed, the third is unstable and slightly perturbed). Bistability occurs because system lies in the lower state until this is no longer possible and then jumps to an high Ras activity correlated state. An interconnected work reproduced the same biological system in a minimal 2D lattice model in order to investigate the relevance of the microdomains clustering in the signaling pathway. It was reported that positive feedback, in the presence of slow diffusion, results in clustering of activated molecules on the plasma membrane and rapid spatial spreading as the front of the cluster of the newly formed membrane bounded Ras-GTP propagates with a constant velocity dependent on the feedback strength (*J.Chem.Phys*, 2009, 130:245102).

Our work, implemented in the software *Smoldyn* (*Phys.Biol*, 2000, 1:137-151), presents a new description of both the above mentioned models by utilizing brownian dynamic stochastic simulations and spatial discretization. It describes how the microdomains clustering of the components of the network can induce bistability.

Interestingly it has recently been found a correlation between the activation levels of components of networks that show ultrasensitivity (such as the Ras/SOS regulatory network) and the Fermi-Dirac statistics generally used to measure the probability of the distribution of the states of a system of

particles (*PNAS*, 2010, **107**:1247-1252). We are carrying out further investigation to test if this kind of description, in which the molecules of the systems can be treated as analogous of fermion gas, could be useful to identify the statistic distribution of the states of a bistable system.

This study has being carried out in the laboratory of professor Piercarlo Fantucci, Department of Biotechnology and Biosciences, University of Milan-Bicocca.

In the last part of the thesis we were finally able to deal with a whole cellular network.

We focused on the heterotrimeric $G_{q/11}$ protein signaling network which is known to be critical for evoking varied physiological functions including sensory perceptions, behavioral and mood regulation, regulation of the immune system activity and inflammation. Dysfunctions of the $G_{q/11}$ network can lead to diseases like cancers and immune system deficiencies. A detailed network topology of the various pathways that emanate from $G_{q/11}$ and that allow signal to flow from the receptor to multiple transcription factors such as AP1, c-fos, c-jun, SRF and CREB has been developed. We have then implemented a multi compartment ordinary differential equation model using the software *Virtual Cell* to analyze signal flow from the plasma membrane through the cytoplasm to the nucleus.

The obtained time courses of activation of key components of the system, such as small GTPases, MAPkinases and transcription factors are in accordance with experimental data provided by the S.Gutkind Lab (Oral and Pharyngeal Cancer Branch, National Institute of Dental and Craniofacial Research, National Institutes of Health, Bethesda, Maryland, USA).

The obtained numerical data will soon be implemented in dynamic graphs to identify the key regulatory motifs of the information processing through the $G_{q/11}$ protein signaling network that are controlling the temporal and spatial

activity of transcription factors such as SRF, c-jun, c-fos, AP1 and CREB.

This study has been carried out in the laboratory of the Professor Ravi Iyengar (Department of Pharmacology and Systems Therapeutics) of the Mount Sinai School of Medicine, New York, USA.

The present thesis allowed us to study four systems each at a different level of complexity, according to the biological question we wanted to answer. The models we proposed constitute another demonstration of the importance of quantitative modeling in biology. This work also suggests that even if all the approaches we used were mathematically different to each other, they share a common methodology. This implies that an underlying correlation runs through all the scales of modeling.

2. Riassunto

Modelli computazionali dell'attività enzimatica di biomolecole a diversi ordini di grandezza: da studi di meccanica quantistica sulla reattività molecolare ad indagini sistemiche di fenomeni cellulari.

Lo scopo di questa tesi è lo sviluppo di modelli computazionali dell'attività enzimatica delle biomolecole a scale diverse di complessità. Indagini parallele sono state condotte a livello quantistico per studiare la reattività di un enzima da un punto di vista elettronico, e ad un livello sistemico mediante tecniche di simulazione per determinare il ruolo degli enzimi nei *network* di reazioni cellulari.

Partendo dal livello inferiore di complessità, il lavoro inizia con due studi computazionali che descrivono il meccanismo molecolare della reazione catalitica, tra la molecola amavadina e metil mercaptoacetato e il coordinamento strutturale tra la proteina angiogenina e lo ione rame. Dato che, in entrambi i casi, sono coinvolti metalli di transizione, sono stati utilizzati i metodi propri della Density Functional Theory (DFT) e della Quantum Theory of Atoms in Molecules (QTAIM) per l'analisi della densità elettronica.

Il primo caso riguarda la descrizione elettronica del meccanismo enzimatico attraverso cui la molecola amavadina, un insolito complesso octa-coordinato di V^{IV} , isolata dai funghi *Amanita muscaria*, catalizza l'ossidazione di alcuni tioli nei corrispondenti disolfuri.

Un lavoro sperimentale di G.da Silva *et al.* Ha proposto un meccanismo a sfera interna per la reazione (*J. Am. Chem. Soc.* 1996, **118**:7568-7573.), ma il meccanismo elettronico non era stato chiarito.

Mediante tecniche DFT, le caratteristiche stereoelettroniche dell'amavadina

allo stato attivo (V^V) e inattivo (V^{IV}) sono state identificate. La formazione del complesso V^V con metil mercaptoacetato (MMA), che è stato scelto come substrato prototipo, è stato quindi caratterizzato sia dal punto di vista termodinamico che cinetico. I risultati rivelano che la protonazione di dell'amavadina (V^V) all'ossigeno carbossilato non direttamente coinvolto nella coordinazione del vanadio, favorisce il legame dell'MMA nella prima sfera di coordinazione del vanadio, in quanto induce la sostituzione dell'ossigeno carbossilato dell'amavadina con quello del substrato e la formazione di un ponte a idrogeno S-H \cdots O. Quest'ultima interazione può promuovere la deprotonazione del gruppo SH e il legame del gruppo tiolato al vanadio. La cinetica e la termodinamica che governano la formazione di intermedi del V^V -MMA sono in accordo con i dati elettrochimici sperimentali e con il ruolo biologico ipotizzato per l'amavadina. È stata infine anche razionalizzata la presenza di un gruppo funzionale estere come un requisito essenziale per l'ossidazione del MMA. Il nostro lavoro costituisce anche una prova indiretta del ruolo del vanadio (nel suo stato attivo d^0) come catalizzatore della protonazione.

Il secondo studio è focalizzato sulle proprietà strutturali e spettroscopiche dei complessi formati dallo ione rame Cu^{2+} e dai frammenti peptidici Ac-PHREN-NH₂, che contengono il putativo dominio di *binding* dell'angiogenina alla cellula, e dai complessi costituiti da Cu^{2+} e dai peptidi della sua variante Ac-PHREN-NH₂. L'analisi delle strutture, delle energie e dei parametri EPR ha permesso di concludere che a pH8 la coordinazione del metallo consta di un atomo di azoto, due gruppi ammidici deprotonati, una molecola d'acqua e un atomo di ossigeno del gruppo COO della catena laterale del glutammato, in accordo con recenti risultati sperimentali [*Dalton Trans*, 2010, **39**:10678]. Risultati ottenuti con tecniche DFT hanno inoltre

consentito di rivelare che la catena laterale del glutammato del peptide Ac-PHREN-NH₂ è coordinato in posizione equatoriale, il che spiega perfettamente gli effetti del legame dello ione Cu²⁺ a questa porzione di angiogenina. Sono stati anche studiati le configurazioni spaziali del sistema mutato Ac-PHRQN-NH₂/Cu²⁺·H₂O. In questo caso, i risultati computazionali hanno portato alla conclusione che la molecola di H₂O è coordinato in posizione equatoriale e che l'atomo di ossigeno del gruppo carbonilico della glutammina è debolmente coordinato in posizione apicale. Utilizzando come riferimento i recenti risultati sperimentali riportati da Bonomo [*Dalton Trans*, 2010, **39**:10678], sono stati impiegati calcoli di meccanica molecolare classica e di dinamica molecolare seguiti da ottimizzazioni DFT per esplorare lo spazio configurazionale del complesso Ac-PHREN-NH₂/Cu²⁺·H₂O.

Entrambi questi studi DFT sono stati svolti presso il laboratorio del professor Piercarlo Fantucci, Dipartimento di Biotecnologie e Bioscienze, Università di Milano-Bicocca.

Questi due studi hanno portato alle seguenti pubblicazioni:

“DFT characterization of key intermediates in thiols oxidation catalyzed by amavadin”, *Dalton Transaction*, 40 (30): 7704-7712, 2011

“Copper coordination to the putative cell binding site of angiogenin. A DFT investigation.”, *Inorg.Chemistry. Accepted*

Acquisite competenze di modellistica a livello molecolare, abbiamo deciso di aggiungere un altro strato di complessità per la nostra indagine e abbiamo deciso di verificare se potesse esistere una correlazione tra la struttura peculiare di una proteina e un effetto biologico a livello cellulare. Per lo studio è stato scelto il sistema biologico fondamentale della pathway

di attivazione di Ras-Sos.

Sulla base di due diversi studi precedenti abbiamo sviluppato un modello che può correlare la bistabilità e lo sviluppo di microdomini a livello della membrana plasmatica in questo *network* di piccole dimensioni. Das *et al.* (*Cell* 2009, **136**:337-351) hanno proposto che l'attivazione di Ras Sos-dipendente nei linfociti provoca una risposta di tipo digitale (*on/off*) che implica la presenza di un *feedback positivo* del meccanismo di regolazione dell'attività di Sos. Il modello deterministico sviluppato dagli autori dimostra che per livelli bassi o alti di Sos vi è un solo stato possibile correlato rispettivamente a livelli bassi o alti di Ras attivo. Per valori intermedi di Sos, si generano tre stati di attività di Ras, di cui due stabili e uno instabile. La bistabilità si verifica allorché per il sistema non è più possibile giacere nello stato inferiore del sistema e quindi, in un determinato momento, salta a un stato di attività superiore. Un lavoro correlato al precedente ha riprodotto lo stesso sistema biologico in un modello minimale a lattice 2D al fine di indagare la rilevanza dei microdomini in questa via di trasduzione del segnale. I risultati hanno rivelato che il *feedback positivo*, in presenza di lenta diffusione, induce il *clustering* delle molecole attivate sulla membrana plasmatica e la rapida diffusione spaziale (*Chem.Phys*, 2009, 130, 245102). Il nostro lavoro, implementato nel software *Smoldyn* (*Phys.Biol*, 1, 137-151, 2000), presenta una nuova descrizione di entrambi i modelli citati, utilizzando simulazioni stocastiche e dinamiche di tipo *browniano* con una risoluzione spaziale ad alta definizione. Il nostro modello descrive come lo sviluppo di microdomini nel sistema di attivazione Ras-Sos possa indurre bistabilità.

E' stata inoltre recentemente trovata una correlazione tra i livelli di attivazione delle componenti di *network* che mostrano ultra-sensibilità (come il *network* Ras-Sos) e la statistica di Fermi-Dirac generalmente utilizzato per

misurare la probabilità di distribuzione degli stati di un sistema di particelle (*PNAS*, 2010, **107**:1247-1252). Stiamo quindi iniziando ulteriori studi per verificare se questo tipo di descrizione, in cui le molecole di un sistema possono essere trattate come analoghi dei fermioni, può essere utile per identificare la distribuzione statistica degli stati di un sistema bistabile. Questo studio è stato svolto nel laboratorio del professor Piercarlo Fantucci, Dipartimento di Biotecnologie e Bioscienze, Università di Milano-Bicocca.

Nell'ultima parte della tesi siamo stati finalmente in grado di occuparci di un'intero *network* di interazioni a livello cellulare.

Ci siamo concentrati sul *pathway* di trasduzione del segnale delle proteine eterotrimeriche $G_{q/11}$, note per essere critiche in diverse funzioni fisiologiche tra cui le percezioni sensoriali e comportamentali, la regolazione dell'umore e della attività del sistema immunitario e i processi infiammatori. Disfunzioni in questo *network* possono portare a malattie come tumori e carenze del sistema immunitario. E' stata inizialmente sviluppata una rete topologica dei vari pathway che scaturiscono da $G_{q/11}$ e che consentono al segnale biologico di fluire dal recettore fino a fattori di trascrizione come AP1, c-fos, c-jun, SRF e CREB. Abbiamo quindi creato un modello computazionale utilizzando equazioni differenziali utilizzando il *software Virtual Cell* per analizzare il flusso del segnale dalla membrana plasmatica attraverso il citoplasma fino al nucleo.

I *time course* di attivazione delle componenti chiave del sistema, come GTPasi, MAPkinases e fattori di trascrizione sono risultati conformi con i dati sperimentali forniti dal laboratorio del prof. S.Gutkind (Oral and Pharyngeal Cancer Branch, National Institute of Dental and Craniofacial Research, National Institutes of Health, Bethesda, Maryland, USA). I dati numerici ottenuti verranno poi implementati nei grafici dinamici per

identificare i *regulatory networks* principali del network che controllano l'attività temporale e spaziale dei fattori di trascrizione come la SRF, c-jun, c-fos, AP1 e CREB.

Questo studio è stato svolto presso il laboratorio del Prof. Ravi Iyengar (Department of Pharmacology and Systems Therapeutics) del Mount Sinai School of Medicine, New York, USA.

Questo lavoro di tesi ci ha permesso di studiare quattro sistemi appartenenti a diversi livelli di complessità. I modelli proposti costituiscono un'altra dimostrazione dell'importanza della modellistica di tipo quantitativo in biologia. Questo lavoro suggerisce anche che, nonostante siano matematicamente molto differenti tra loro, gli approcci utilizzati condividono metodologie comuni implicando l'esistenza di una correlazione di fondo tra tutte le scale di *modeling*.

3. Introduction

3.1. The role of modeling in cell biology

"All right," said Deep Thought. "The Answer to the Great Question . . . "
"Yes . . . !"
"Of Life, the Universe and Everything . . . " said Deep Thought.
"Yes . . . !"
"Is . . . " said Deep Thought, and paused.
"Yes . . . !"
"Is . . . "
"Yes . . . !!! . . . ?"
"Forty-two," said Deep Thought, with infinite majesty and calm.
(...) "Forty-two!" yelled Loonquawl. "Is that all you've got to show for seven and a half million years'
work?"
"I checked it very thoroughly," said the computer, "and that quite definitely is the answer.
I think the problem, to be quite honest with you, is that you've never actually known what the question is."

Douglas A. Adams - The hitchhiker's guide to the galaxy [*]

In the last years the number of studies combining experiments with quantitative modeling in biology has been constantly arising.

Even if most of the chemical reactions that occur in the cell simply involve a single enzyme catalyzing a transition in a substrate so that their evolution during time can be plotted with just a couple of Michaelis-Menten constants, both physics and chemistry has shown us, however, that even when but a few players interact in nonlinear ways, complexity and large-scale phenomena soon arise. Nevertheless this does not prevent us to describe them by more complicated mathematical equations or computer code and it doesn't necessarily prevent their predictability.

The model of a system can be considered successful if observed behaviors emerge from qualitative or quantitative analysis of the equations that constitute it. So, "mathematics can be thought of as a crutch for our

intuition, to help us bridge the gap between what we can see and what we can think about [2]”.

But, when biological events are taken into consideration, complexity rapidly increases and understanding how cellular-scale behaviors arise from molecular actions becomes very difficult due to the large number of many different kinds of molecules all interacting in complex networks. Another difficulty lies in the scale of the cell biological systems which aren't microscopic but they are not macroscopic either.

The inherent redundancy and heterogeneity of the molecular machines makes cell modeling very challenging but also unavoidable because of the huge amounts of experimental data that have been produced in the last few decades.

Since these data are often quantitative and statistical data a mere qualitative assembly and description of them is not longer useful. It's here that quantitative modeling surges at its fundamental role in generating hypotheses and a natural endpoint for the experimental efforts.

According to Mogliner et al. [2] there are three fundamental questions that a modeler should ask him/herself before getting into action:

“What exactly does modeling add to experimental studies?

What is the appropriate model scale?

What are the appropriate modeling methods?”

The first consideration is that the model should not be just a mere elaboration of the experimental data or the mathematical transduction of a biological issue but it has to provide a new insight on the scientific question. The model should at least recapitulate the phenomenon of interest so that it can indicate if all of the necessary players and interactions have been

identified or not. Moreover, the model should help to investigate those aspects and characteristic that cannot be detected by experimental data.

The choice of the modeling methods depends on the purpose and the scale of the model. The latter is the most critical issue in developing a model because the *scale* is strongly correlated to the *biological question* we need to answer. This is not an assertion as trivial as it can seem. To avoid sharing as the same experience as the hyper-intelligent scientists in Adam's novel who only after ages disappointingly find out that model they built was inadequate to answer their question [1], it is fundamental that the scientific question motivating the model strongly relies on a satisfactory amount and quantity of data and an appropriate methodology. The process of modeling by itself forces the modeler to ask many important questions: have all the correct components be included? Are all the different parts important in the same way? Which is the best approach to simulate this phenomenon?

Choosing between a fine-grained and a coarse-grained model involves several trade-offs: a coarser-grained model will often rely on extensive prior intuition about the cellular phenomenon while a finer-grained model could require more detailed prior information about the properties of the individual components.

From a mathematical point of view the simplest models describes the functional dependence of one observable from another and in this case algebraic equations are sufficient.

The next level of mathematical complexity could be considered the deterministic temporal dynamics described by ordinary differential equations (ODEs). This formalization is often use to model cell signaling and metabolic pathways in cases when the cell can be considered as a well-stirred reactor and when stochastic effects can be neglected.

A unique feature of cells is the utilization of spatial separation of

components and interactions. Partial differential equations (PDEs) can be used to model spatially heterogeneous and compartmentalized dynamics.

The role of quantitative modeling in biology as a complementary instrument of biological discovery will reasonably continue to increase. Building a good model depends too much on intuition, on the rare abilities to ask the right question and to sense mathematical order behind messy facts, on tricky timing (not too early, when absence of data leaves a model unsubstantiated, not too late, when everything is clear without a model), and on hard, long thinking that makes modeling so painful, but also so much fun.

3.2 Aim of the thesis

In this work, we developed four different model in relation to as many biological question. We started from “simple” models that involve just on or two molecules, in order to identify their catalytic and binding mechanism. But the structure of a molecule can induce phenomena at a wider level involving the space distribution of other molecules in the cell. This was the object of the second part of our investigation. Finally we decide to study which structure can arise from more complex systems and we focused on the cell behavior.

The models we finally propose constitute another demonstration of the importance of quantitative modeling in biology. Our work also suggests that even if all the approaches we used were mathematically different to each other, they share a common methodology. This implies that an underlying correlation runs through all the scales of modeling.

3.3 References

- *) The hitchhiker's guide to the galaxy, Douglas Adams, 1979
- 2) Mogilner *et al.*, *Developmental Cell*, 2006, **11**:279-287

4. Quantum mechanic studies of reactivity mechanism of bioinorganic molecules

“The principle of generating small amounts of finite improbability (...) were often used to break the ice at parties by making all the molecules in the hostess's undergarments leap simultaneously one foot to the left, in accordance with the Theory of Indeterminacy.

Many respectable physicists said that they weren't going to stand for this — partly because it was a debasement of science, but mostly because they didn't get invited to those sort of parties.”

The hitchhiker's guide to the galaxy - Douglas Adams,[1]

In the first part of this thesis it was chosen to study biological systems that lie in the lower level of complexity, i.e. the interaction between two molecules or a molecule and an atom.

Two systems were selected: amavadin, a bioinorganic component of some species of *fungi* and angiogenin, a human protein that is involved in angiogenesis. The first is interesting because of its peculiar structure and the small amount of literature data on its reaction mechanism with thiols, while the second has been chosen in order to investigate its role in binding copper in the human-life threatening angiogenetic process.

In order to study these systems, Quantum Chemistry methods, in particular methods derived from the Density Functional Theory (DFT), were used as it is explained in the following section.

4.1 Methods of Quantum Chemistry

The most evident characteristic of the two selected molecules is that they interact with a transition metal: amavadin's inner core is an atom of vanadium (V), while angiogenin coordinates an atom of copper (Cu). Because of this, methods proper of classic Molecular Mechanics as *force fields*, are not adequate for a good description of this kind of systems in which stereoelectronic effect must be taken into account.

So, it was decided to employ the methods of Quantum Chemistry that explicitly treat electrons and allow to study in detail the effects that are the basis of coordination between the transition metal and the molecule. Quantum Mechanics (QM) is a physical theory formulated in the first half of the twentieth in order to describe the behavior of the matter at the microscopic level. It fundamentally lies onto two assumptions. The first is the de Broglie's hypothesis of *wave-particle duality*, in which matter has a wavelike nature during the temporal evolution of the system and a particle quantum-type nature, at the time of measurement; the second is the Heisenberg's *uncertainty principle*, according to which there are pairs of variables (those that are not compatible with each other), such as the position and the momentum of a particle, whose values can not be known simultaneously with arbitrary precision, regardless on the accuracy of measurements. Quantum uncertainty can only predict the probability of measuring certain values of conjugate variables at the time of the experiment.

The fundamental difference between classical and quantum approaches is expressed in the discussion of the location of a particle: while the Newtonian mechanics associated with a particle trajectory $x(t)$, $y(t)$, $z(t)$ are time

dependent, quantum mechanics relates to the particle wave function $\psi(x, y, z)$, such that its squared module $|\Psi(x, y, z)|^2$ gives the probability distribution of finding the particle at the point with coordinates (x, y, z) .

Quantum chemistry is a discipline of theoretical chemistry that studies the properties of atomic and molecular *ab-initio* on the basis of the *Schrödinger equation*. This fundamental equation is used to calculate the wave function of a molecular system from the physical constants (speed of light, Planck's constant, the electron and the proton charge):

$$H\Psi = E\Psi$$

where H is the Hamiltonian operator associated to the total energy of the system, E is the energy of the system and ψ is the wave function.

For a molecule consisting of M nuclei, each of charge $+Z_{\alpha}$, and N electrons, the equation takes the following form:

$$\left[\left(\frac{\hbar^2}{8\pi^2 m} \right) \sum \nabla^2 - \sum \sum \frac{Z_{\alpha} e^2}{r_{\alpha i}} + \sum \sum \frac{e^2}{r_{ij}} \right] \psi = E\psi$$

where $-(\hbar^2/8\pi^2 m) \nabla^2$ is the kinetic energy operator, $Z_{\alpha} e^2/r_{\alpha i}$ is the attraction between the nucleus and the electron α and e^2/r_{ij} is the repulsion between the electron i and electron j .

Quantum chemistry explicitly treats a system of electrons, so the methods used in quantum molecular modeling start from a given nuclear geometry and determine the best distribution of the electrons in the potential generated

by the nuclei. These methods can be divided into two classes: the *ab initio* methods, such as the Hartree-Fock method, and the methods that are derived from the *density functional theory* (DFT), that directly solve the Schrödinger equation for the system, without simplifications due to experimentally obtained parameters. The *semiempirical* methods, such as the Huckel method, use parameters experimentally obtained to solve the Schrödinger equation in order to avoid time consuming calculations.

4.1.1. Density Functional Theory methods

The Density Functional Theory (DFT) is an approach to the electronic structure of atoms and molecules that has experienced considerable growth in recent years. The DFT's purpose is to calculate the total electronic energy of the system starting from the electron density. The basic principle is that there is a relationship between the energy and the electron density of the system, ρ . Although this is not a concept born in the last few years, as similar assumptions were made since the '20s, it was only with Hohenberg and Kohn, in 1964, that this relationship was uniquely defined by the expression according to which the energy is a functional of $\rho(r)$. A functional allows a function to be reduced to a scalar and is therefore usually written as:

$$Q[f(r)] = \int f(r) dr$$

the function $f(r)$ is generally dependent on other functions. In the case of DFT, the function is the electron density. In the DFT, the energy functional is written as a sum of two terms:

$$F[\rho(r)] = \int V_{\text{ext}}(r)\rho(r)dr + F[\rho(r)]$$

where the first term expresses the interaction of electrons with an external potential (usually due to coulombic interactions between the nuclei), $F[\rho(r)]$ is the sum of the kinetic energy of the electrons and the contribution of interelectronic interactions. The minimum energy (i.e. the best solution) corresponds exactly to the electron density.

Electron density is constrained to the number of electrons, N :

$$N = \int \rho(r)dr$$

To minimize energy a Lagrange multiplier ($-\mu$) is introduced as a constraint and that gives to:

$$\frac{\delta}{\delta \rho(r)} \left[E[\rho(r)] - \mu \int \rho(r)dr \right] = 0$$

from here you can write:

$$\left(\frac{\delta E[\rho(r)]}{\delta \rho(r)} \right)_{V_{\text{ext}}} = \mu$$

that is the DFT equivalent of the *Schroedinger equation*. V_{ext} indicates that it is under constant external potential conditions (nuclear positions fixed).

It is interesting to note that the Lagrange multiplier, μ , can be identified with the chemical potential of an electron cloud around the nuclei. This is somewhat related to the electronegativity:

$$-\chi = \mu = \left(\frac{\partial E}{\partial N} \right)_{V_{\text{ext}}}$$

In 1965, Kohn and Sham suggested a practical way to solve the Kohn and Hohnberg theoreme for a set of interacting electrons. The inherent difficulty of the theorem is that you don't know what function $F[\rho(r)]$ is. The authors suggested that $F[\rho(r)]$ could be approximated as the sum of three terms:

$$F[\rho(r)] = E_{KE}[\rho(r)] + E_H[\rho(r)] + E_{XC}[\rho(r)]$$

where $E_{KE}[\rho(r)]$ is the kinetic energy, $E_H[\rho(r)]$ is the Coulomb electrostatic energy and $E_{XC}[\rho(r)]$ contains the contribution of the exchange and correlation. It is important to note that $E_{KE}[\rho(r)]$ is defined as the kinetic energy of a system of not-interacting electrons with the same density $\rho(r)$ of the real system; the second term is also known as electrostatic energy Hartree.

The complete expression of energy for a N-electron system within the framework of Kohn-Sham is:

$$E[\rho(r)] = \sum_{i=1}^N \int \psi_i(r) \left(-\frac{\nabla^2}{2} \right) \psi_i(r) dr + \frac{1}{2} \iint \frac{\rho(r_1)\rho(r_2)}{|r_1 - r_2|} dr_1 dr_2 + E_{XC}[\rho(r)] - \sum_{A=1}^M \int \frac{Z_A}{r - R_A} \rho(r) dr$$

Kohn and Sham wrote the density of the system as the sum of the squared modules with a set of orthonormal monoelctronic orbitals:

$$\rho(r) = \sum_{i=1}^N |\psi_i(r)|^2$$

to solve the Kohn-Sham equation an electron density, calculated from the electronic wave function at the initial geometry, is iteratively inserted. The expression of the density is then optimized until convergence is reached.

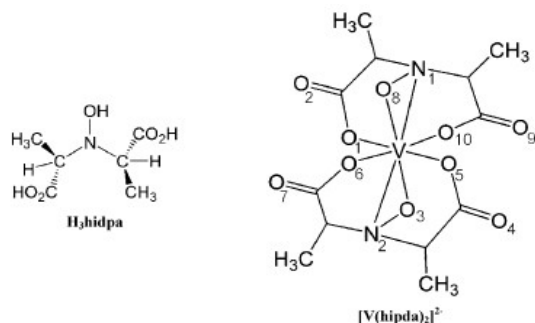
There are several features that help solve the Kohn-Sham equation in dependence of the characteristic of the systems. Computational details for both the amavdin and the angiogenin system are given in the next sections.

4.2. DFT characterization of key intermediates in thiols oxidation catalyzed by amavadin

4.2.1 The biological question

For its ubiquitous distribution in the entire biosphere and its highly amphoteric oxidative character, vanadium has been found to have so much relevance in biological systems that its role as an essential metal in the first steps of evolution has recently been hypothesized [1]. Representing 0.014% of the lithosphere and the second-most abundant transition metal in seawater, with a concentration of 30 nM, [2] vanadium is widely available to living organisms. Moreover, the ease of shifting between three oxidation states (+3, +4, +5) of biological relevance further emphasizes its capability to serve as a redox shuttle.

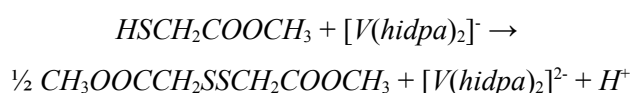
Some species of the genus *Amanita*, e.g. *A.muscaria*, *A.muscaria regalis*, *A.velatipes*, show concentrations of vanadium of up to 400 mg kg⁻¹ in the form of a natural vanadium(IV) compound, called amavadin [3,4]. It is well known that the specific chemical environment of the vanadium centers in natural vanadium complexes and proteins plays a fundamental role in tuning the metal biological activity. In this scenario the very peculiar structure of amavadin has raised growing interest from both a chemical and biological point of view. Amavadin is a 2 : 1 octa-coordinated complex between V^{IV} and *S,S*-2-2'-hydroxyimino-dipropionic acid HON{CH(CH₃)COOH}(H₃hidpa) in its deprotonated form, and metal coordination involves the carboxylic and hydroxyamino groups of the ligand [5]:



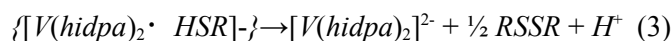
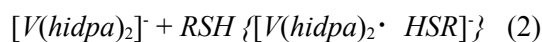
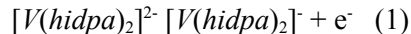
Remarkably, before isolating it in the amavadin crystal structure, the H₃hidpa ligand was completely unknown. The crystallographic characterization of amavadin [6] ([V(hidpa)₂]²⁻ hereafter) revealed that no vanadyl group [VO]²⁺ is present, as opposed to the precedent hypothesis [7]. The complex displays an intriguing *non-oxo* octacoordinated V^{IV} ion with a distorted dodecahedral {VO₆N₂} coordination, with two η² N–O of the oxyimino group and four unidentate carboxylate groups. The relative orientation of the two N–O leads to the D and K chirality of the metal center. In fact, spectroscopic studies [6] have revealed that isolated amavadin consists of an equimolar racemic mixture of two stereoisomers, D and K.

For its reversible one-electron redox properties, amavadin can act as a catalyst for thiol oxidation [8], peroxidative halogenation, hydroxylation and oxygenation of alkyl and aromatic substrates [9]. Recently Guedes da Silva *et al.* reported that amavadin can also convert methane into acetic acid [10] and ethane into propionic and acetic acid [11]. In light of these findings, it is plausible to define amavadin as a sort of proto-enzyme even though its real function has not yet been identified[2]. Hubretgste *et al.* [12] demonstrated that the carboxylate groups are essential for the stability of amavadin and consequently an outer rather than inner sphere catalytic mechanism for a series of electrocatalytic transfer reactions was initially hypothesized.

Notwithstanding, prompted by a work by Tackerey and collaborators [13], Guedes da Silva *et al.*[14] gave experimental evidence of an inner sphere mechanism in the oxidation of thiols (RSH) to the corresponding disulfides (RSSR) by amavadin. The latter reaction, which could be involved in the protective/defensive system of the mushroom since it could regenerate broken disulfides bridges in protein fibers of damaged tissues [14], involves the one electron oxidized V^V form of amavadin with the formation of the corresponding disulfide.



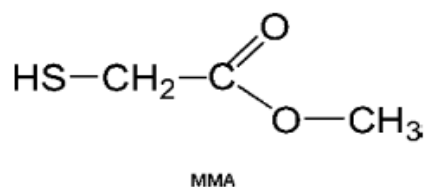
In particular, the following reaction steps were proposed on the basis of cyclic voltammetry measurements:



This redox catalysis process occurs through a Michaelis–Menten kinetic mechanism, as revealed by saturation of the catalytic efficiency upon increasing substrate concentration. The active catalytic species is the d⁰ V^V amavadin and the transient formation of an intermediate species [V(hidpa)₂ · HSR]-generated by coordination of the oxidized vanadium complex to the thiolic substrate has been observed. In particular, Guedes da Silva *et al.* found a rate constant for the forward reaction (2) of 1.2x 10³ M⁻¹s⁻¹, and estimated a half life of 0.3 s for the [V(hidpa)₂ · HSR]⁻ intermediate, when mercaptoacetic or mercaptopropionic acid were used as substrates.

Quantum chemical approaches are useful tools in the characterization of the stereoelectronic properties of coordination compounds. In fact, the electronic structure of amavadin in its d^1 reduced and oxidized d^0 forms was investigated at DFT level in a few previous theoretical works. Armstrong and coworkers [15] presented a $X\alpha$ investigation in order to elucidate character of the frontier MOs, showing the similarity of the $d^{(1)}x^2-y^2d^{yz}d^{xz}d^{yz}d^{z^2}$ d-orbital ordering of $[V(\text{hidpa})_2]^{2-}$ with that of oxovanadium(IV) VO^{2+} complexes, as $[\text{VOCl}_4]^{2-}$. In light of these results, the electronic spectrum of the reduced amavadin form was computed and compared with the experimental one. Remenyi and co-workers [16] and Ooms and co-workers [17] dealt with the computation of EPR and ^{51}V NMR parameters of the reduced form of amavadin, while in a more recent paper Geethalakshmi [18] and co-workers investigated the ^{51}V NMR parameters of the oxidized form.

To carry out the present investigation Density Functional Theory (DFT) has been used to study the reactivity of amavadin with methyl mercaptoacetate (MMA), which is known to undergo oxidation to the corresponding disulfide, according to the mechanism proposed by Guedes da Silva *et al.* [14]. Then, the electronic structure and the bonding properties of key species involved in the catalytic process were investigated by means of the Quantum Theory of Atom in Molecules approach (QTAIM), which was never used before to study the stereoelectronic features of amavadin.



The present study also fits into the line of investigations about the controversial vanadate reduction by thiols. A recent work by Crans *et al.* [19] shows that both V^{IV} and V^V can form complexes with thiols and that the redox chemistry that can occur is governed by different parameters such as pH, solute and vanadate-concentrations, as well as the presence of other ligands. The present investigation has thus taken into account the possibility of forming a methylmercaptoacetate–amavadin adduct through V^V –S bond making. The role of an ancillary (with respect to the thiol) functional group, such as an ester, has also been elucidated.

4.2.2. Computational details

Computations were performed using the pure Generalized Gradient Approximation (GGA) B-P86 DFT functional [20] and the Resolution of Identity [21] technique, as implemented in the TURBOMOLE [22] suite of programs. Basis sets of triple-zeta plus polarization split valence quality (TZVP hereafter) [23] were adopted for all atoms in the complexes. The DFT grid-size was set to the standard m3 value. The employed theory level has turned out [24,25] to be well suited for DFT investigations regarding the reactivity properties of V^V complexes, which are closely related to amavadin, such as the cofactor of vanadium haloperoxidase (VHPO). Solvent effects were taken into account using the COSMO(Conductor-like Screening Model; $\epsilon = 80.0$) approach [26-28]. The pseudo-Newton–Raphson procedure has been adopted for the search of transition states (TS). Schematically, it entails the structure optimization (SO) of a guess of the true TS. During this SO, those degrees of freedom which are supposed to vary most along the reaction coordinate (RC) are kept frozen; this step allows all the structure but the reactive core to reach a potential energy minimum. After calculating the

Hessian matrix associated to the optimized guess TS, (the presence of a negative eigenvalue corresponding to the motion along the RC ensures about the nature of the found saddlepoint) an eigenvector-following procedure is adopted for locating the maximum energy point along the RC, *i.e.*, the TS. The rototranslation partition function has been evaluated for some key points of the catalytic cycle of thiol oxidation, in order to correct E values by means of entropy and Zero Point Vibration Energy contributions, so as to obtain an approximate Gibbs free energy (G) for crucial species. QTAIM analysis of the electron density was implemented using AIMPAC [29] and SF_ESI³⁰ codes. The Ω atomic net charges are computed with an accuracy on the numerical integration such that the atomic lagrangian is less than of 10^{-4} . Calculations have been carried out only on the D stereoisomer of amavadin, so in the following the symbol D will be omitted for nomenclature simplification.

4.2.3. Results and discussions

To shed light on the structural features of key intermediate species formed in the inner sphere mechanism of thiol oxidation experimentally studied by Guedes da Silva *et al.*, [14] the coordination mode of a prototypical substrate (MMA) to the vanadium atom of amavadin has been studied using DFT. In fact, experimental evidence showed [1] that the presence in the substrate molecule of a carboxylic derivative two or three bonds away from the sulfhydryl group is necessary for catalysis to take place. Moreover, both H-S-R-COOR' thiols with R' corresponding to carboxylate or ester groups undergo oxidation, indicating that the hydroxylic group of the carboxyl moiety of the substrate does not play a fundamental role in the reaction.

According to this observation, MMA rather than mercaptoacetic acid has been chosen as the substrate in our calculations. To shed light on the reaction mechanism, the possibility of a protonation event on amavadin has been also taken into account, since catalysis takes place in a KCl–H₂O electrolyte solution [14]. Therefore, both protonated and non-protonated forms of oxidized amavadin have been considered.

4.2.3.1. Geometry and electronic structure of reduced and oxidized forms of amavadin

The V^{IV} [V(hidpa)₂]²⁻ structure has been optimized starting from its X-ray crystallographic structure, while the V^V [V(hidpa)₂]⁻ structure has been optimized starting from the optimized [V(hidpa)₂]²⁻ atomic coordinates. The vanadium coordination in [V(hidpa)₂]²⁻ and [V(hidpa)₂]⁻, as well as DFT optimized bond distances, δ(A,B) delocalization indexes, and QTAIM atomic net charges are shown in Fig. 1.

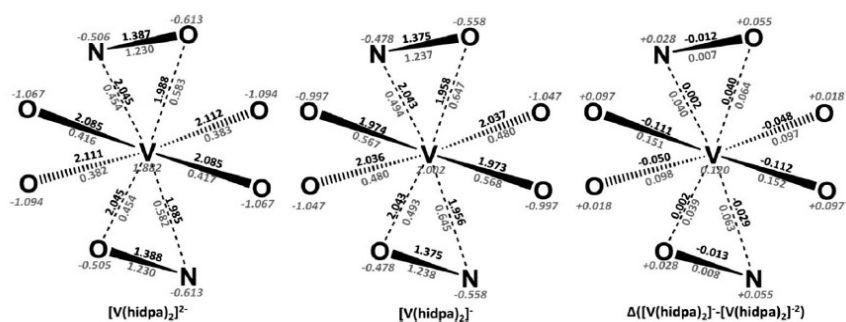


Fig. 1 From left to right: [V(hidpa)₂]²⁻ and [V(hidpa)₂]⁻ {VO₆N₂} first vanadium coordination sphere optimized bond distances (in black, in Å), delocalization indexes (grey) and atomic net charges (grey, italics). In the last picture differences between the properties of the V^V–V^{IV} species computed are reported. [V(hidpa)₂]²⁻ X-ray bond distances: V–NO 1.999(8)Å, 1.982(8)Å; V–ON 1.940(7)Å, 1.956(7)Å; V–O 2.070(8)Å, 2.028(7)Å; 2.028(9)Å, 2.042(8)Å; 2.070(8)Å (Ca²⁺ counterions, see ref. 34); [V(hidpa)₂]⁻ X-ray bond distances: V–NO 1.999Å, 2.0Å ; V–ON 1.940Å, 1.923Å ; V–O 1.993Å, 1.977Å; 1.972Å,

1.959Å. (PPh₄⁺counterions, see ref. 15)

On average, the [V(hidpa)₂]²⁻ optimized V–O and V–N bond distances are overestimated by 0.030Å, with respect to X-ray diffraction values [35]. The same comparison made for optimized [V(hidpa)₂]⁻ bond distances revealed a similar average error (0.034Å). The quality of the optimized geometry pentameters obtained using BP86 functional does not change significantly adopting a different DFT functional.

The oxidation of [V(hidpa)₂]²⁻ to [V(hidpa)₂]⁻ releases 13.7 kcal mol⁻¹ but does not modify significantly the structure of the complex, in agreement with experimental data. However, a contraction of the V–O(1) and V–O(5) bonds (5.3–5.4%) and of the V–O(6) and V–O(10) bonds (3.7%) occur, whereas the V–O(3) and V–O(8) bonds show a modest contraction (1.5%) and the V–N bond a minimal shrinkage (0.1%).

The effect of the one-electron oxidation on the electronic structure of [V(hidpa)₂]²⁻ depends on which property is discussed.

The increase of the atomic net charges is essentially distributed over the whole system. The total increase of the atomic net charges of the V atom and its {VO₆N₂} first coordination sphere is 0.52 (see Fig. 1, right), where the V positive charge increases by 0.12 e. However, the spin population of [V(hidpa)₂]²⁻, which is completely localized on the V atom (a-b = 1.011 e), disappears in the closed shell [V(hidpa)₂]⁻. This last point is also evidenced in the analysis of the population of the [V(hidpa)₂]²⁻ Singly Occupied Molecular Orbital (SOMO, see Fig. 2).

The total V population of the SOMO is 0.8 electrons completely due to the d-orbitals contributions. It is interesting to note that the largest contribution is due to the d^{x²-y²} (0.35 electrons) followed by the d^{yz} contribution (0.29 electrons).

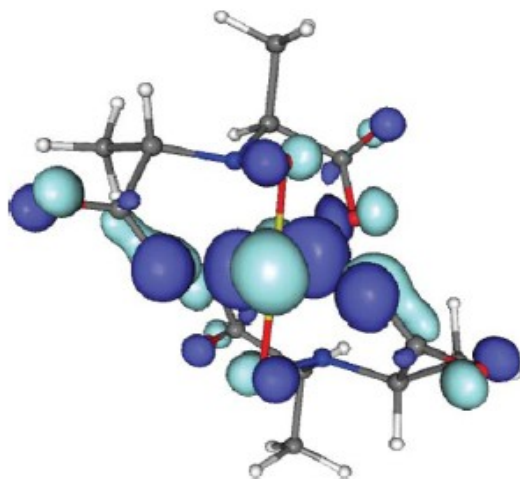


Fig. 2 Isosurface (0.03) plot of the highest Singly Occupied Molecular Orbital (SOMO) of [V(hidpa)]²⁻.

4.2.3.2. DFT optimization of the protonated form of [V(hidpa)₂]⁻

The kinetics data [14] reported by Guedes da Silva *et al.* show that no pH effects were detected on the electrocatalytic activity of amavadin. The authors of such experimental work proposed that this behavior could depend on the high stability constants of the amavadin complex ($>10^{21}$), implying that the catalyst remains undissociated even at very low pH values. This observation is apparently in contrast with the proposed inner sphere mechanism, since if the chelate-V bonds are too strong to be cleaved, formation of a vacant site for MMA coordination would be prevented. In addition, it must be also noted that substitution of one of the carboxylate groups of the hidpa ligand with the ester group of MMA would not be energetically favorable. In this scenario, it is conceivable that protonation of one vanadium ligand in amavadin can take place and weaken one of the V–X bonds, thus favoring the coordination of the substrate to the active redox

form of the catalyst. Prompted by this observation, the $[\text{V}^{\text{V}}(\text{hidpa})_2]^-$ structure has been subjected to a systematic sampling of all the possible protonation sites, with the aim of locating the most basic sites (according to the Brønsted–Lowry definition) of the V complex. Each oxygen atom not included in the first V coordination sphere has been initially protonated considering the two different possible orientations of the H atom with respect to V (in structures A and B the H atom points toward or in opposition to the V atom, respectively). Similarly, both “up/down” orientations (relative to the equatorial plane of the complex) have been considered for protonation at those O atoms which are ligated to V. The protonated forms of amavadin will be referred to as $[\text{V}(\text{hidpa})_2\text{H}]$. Protonation at the N side is not expected to occur, since the negatively charged oxygen atom of the *N*-oxide group should afford relatively more stable structures. Anyway a DFT optimization has been performed to check the reliability of our assumption: the N–H isomer turned out to be highly unstable (more than 34 kcal mol⁻¹ than the most stable O–H structure). Relative stability data for all protonated forms investigated, as well as the V–O and V–N bond distances for the four more stable isomers, are reported in Table 1.

It turned out that amavadin protonation is always favored at the oxygen atom not directly bound to V. In fact, the four structures protonated respectively at O(2), O(4), O(7) and O(9) are essentially isoenergetic, with energy differences that span in the narrow range of 0–0.3 kcal mol⁻¹. The other protonated structures differ in energy by at least 5.1 kcal mol⁻¹ from this group. Note that in most cases, the starting point structure of type A and B converged to the same minimum.

The structure of the two $[\text{V}(\text{hidpa})_2\text{H}]$ isomers, which differ in energy by 5.1 kcal mol⁻¹, is shown in Fig. 3.

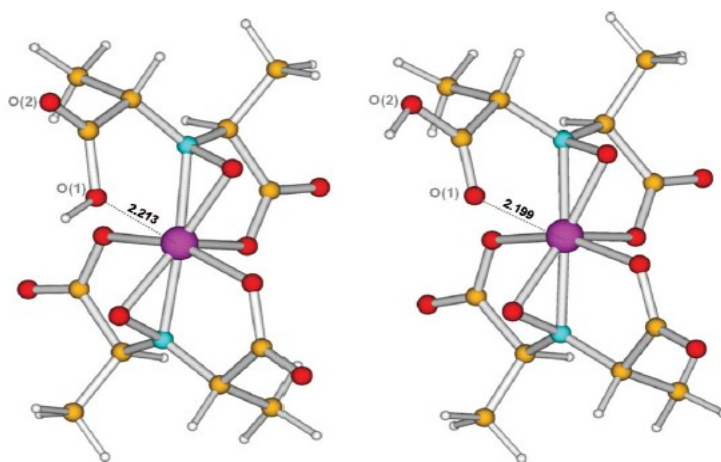


Fig. 3 Optimized geometries of VV [V(hidpa)₂H] isomers in which O(1) (left) or O(2) (right) is protonated. The V–O distances are in Å^o. Vanadium is in yellow, nitrogen in cyan, oxygen in red, carbon in orange and hydrogen in white.

Clearly the influence on [V(hidpa)₂H] geometry of protonation at O(2) is lower than that at O(1). In particular, the {VO₆N₂} coordination is perturbed by the lengthening of the V–O(1) bond distances (+0.239 Å and, +0.223 Å, with respect to [V(hidpa)₂]⁻, respectively). It is worth noting that except for the atoms directly involved, protonation does not sensibly modify the molecule geometry, since just in one out of the sixteen optimizations that have been carried out a significant change occurs. In this particular case, the V–O(10) bond is elongated by 58% and the structure lies 11.4 kcal mol⁻¹ higher in energy compared to the most stable one. The QTAIM analysis of the electron density of the structure protonated at O(2) shows that the effect of protonation on the {VO₆N₂} moiety is essentially localized on the V–O(1) bond. In particular, the V–O(1) bond critical point [38] is still found but the $\delta(V-O(1))$ index is almost halved (-0.31 electron couples, -44%), highlighting the strong weakening but not the dissociation of the V–O(1)

collected in Table 2. The van der Waals complex formed between the reactant molecules before the adduct $[V(\text{hidpa})_2\text{H}] \cdot \text{MMA}$ has been formed, features an hydrogen bond between the protonated carboxylate of $[V(\text{hidpa})_2\text{H}]$ and the ester group of MMA.

In particular, the carbonyl oxygen of MMA becomes part of the $\{VO_6N_2\}$ coordination sphere, with a $V\text{-O}(\text{MMA})$ bond distance of 2.150Å; the hydrogen atom of the protonated carboxylate O(2) forms a hydrogen bond with O(6); the hydrogen atom of the thiol group forms a hydrogen bond with the carbonyl oxygen of the protonated carboxylate group (H–O distance = 2.061Å). Other $[V(\text{hidpa})_2\text{H} \cdot \text{MMA}]$ isomers, featuring a thiolate group and neutral carboxylic groups of hidpa, are at least 8.2 kcal mol⁻¹ higher in energy.

In order to also characterize kinetically the found adduct, a transition state (TS) along the reaction coordinate of formation has been searched. The located TS structure, reported in Table 2, is approximately quite similar to the $[V(\text{hidpa})_2\text{H} \cdot \text{MMA}]$ found and, therefore, the same sketch has been used to describe both structures (of course evidencing that the coordination of MMA to the catalyst occurs after the formation of the H-bond between the protonated O(1) and O(6), since such an interaction is detected already in the TS structure. The activation barrier computed with respect to the $[V(\text{hidpa})_2\text{H}] \cdot \text{MMA}$ van der Waals complex is 20.0 kcal mol⁻¹ (21.2 when entropy contribution are included), a value which is indicative of the kinetic accessibility of formation reaction. Globally, both thermodynamics and kinetics of the adduct formation are consistent with a relatively facile reaction and therefore support strongly the possibility of an unprecedented example of inner sphere electrocatalysis turn overed by the amavadin.

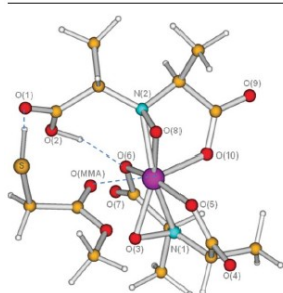
Another interesting point regarding the most stable adducts found by DFT is that their second coordination sphere of vanadium turns out to be

characterized by fluxionality of outer intramolecular hydrogen bonds. Indeed switching from the bottom left adduct in Scheme 1 to the right bottom one does not imply any energy cost, from a thermodynamical standpoint.

Turning the attention to less stable isomeric forms which have been located concerning the substrate–catalyst adduct, it is worth focusing on the 5.0 and 15.2 kcal mol⁻¹ species. Both feature S-to-V coordination, the former being stabilized by an intramolecular H bond between ester group on MMA and a noncoordinated carboxylic group whereas the latter is characterized by the formation of an intramolecular 5-membered chelating ring, through the formation of a second V–O(MMA) bond. It is fair to underline that the relative stability which characterizes these two isomers is likely to be overestimated in favor of the monocoordinated species due to the use of an implicit treatment of solvent effects. Indeed the 15.2 kcal mol⁻¹ form, which features two free carboxylic groups, is expected to establish a number of H bond interactions with water solvent molecules.

The effect of the adduct formation on the electronic structure of the substrate and amavadin by analyzing the QTAIM parameters of the atoms and bonds involved in this step of the reaction were taken into account. As a general remark, the formation of the intermediate has no dramatic effects on the electronic structure of amavadin, although some small but significant effects on MMA are highlighted.

Table 2 Structure of the [V(hidpa),H]-MMA adduct (product) and transition state (TS) which is associated with its formation by starting from reactants (*i.e.* an intermolecular van der Waals complex which is stabilised by a hydrogen bond between MMA and protonated amavadin). Vanadium is shown in yellow, nitrogen in cyan, oxygen in red, carbon in orange, hydrogen in white, sulfur in green. The interactions between MMA and amavadin and the new intramolecular H bond are indicated by a dashed line. Distances are in Å

	TS of ([V(hidpa),H]-MMA) formation	[V(hidpa),H]-MMA	
	V-O(MMA)	2.623	2.150
	V-O(3)	1.958	1.958
	V-O(6)	2.005	2.006
	V-O(8)	1.932	1.932
	V-O(10)	1.963	1.963
	V-O(5)	1.932	1.932
	V-O(2)-H	3.615	3.614
	V-N(1)	2.045	2.045
	V-N(2)	2.129	2.129

The formation of the H bond between the hydrogen of the SH group and the oxygen of the carboxylate ($H\cdots O$ distance 1.656 Å) and the corresponding weakening of the S–H bond (the value of $\delta(S,H)$ decreases from 1.042 in free MMA to 0.980 in the intermediate) is able to make the S atom slightly more nucleophilic. Indeed, the initial low but positive net charge of S in free MMA (+0.0167) becomes negative (-0.011), with an increase of the negative charge by 0.027 electrons. This fact is also well evidenced by the shape of the HOMO of the intermediate, that highlights a significant contribution of the sulfur atomic orbitals (see Fig. 5, bottom view).

The V–O(MMA) interaction is weak, with a $\delta(V-X)$ delocalization index of only 0.28 electron pairs, which is less than half of the value for other V–O interactions with the carboxylate of the hidpa ligands.

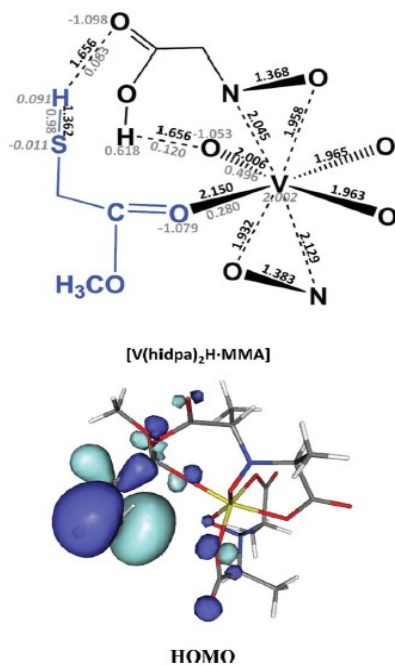


Fig. 5 Top: schematic view of the [V(hidpa)₂H · MMA] intermediate with {VO₆N₂} vanadium coordination sphere (in black) optimized bond distances (in black, in Å), delocalization indexes (grey) and atomic net charges (grey, italics). MMA is highlighted in blue. Bottom: iso-surface (0.03) plot of the [V(hidpa)₂H · MMA] HOMO.

4.2.3.4. Characterization of reaction intermediates formed when MMA reacts with [V(hidpa)₂]⁻

The stereoelectronic characterization of key intermediates formed in the reaction between amavadin [V(hidpa)₂]⁻ and MMA is crucial to cast light on this unprecedented inner-sphere oxidation mechanism, operated by an otherwise outer-sphere performing catalyst.

As shown in the previous section, all the plausible coordination of MMA to [V(hidpa)₂H] have been considered, and the different isomers characterized

by DFT, as well as their relative energies (Scheme 1).

In light of the results published by Crans *et al.*, [19] which demonstrated that a mercapto-aminoacid like cysteine can bind to V^V in a non-reductive way through the thiol group, initially the possibility has been considered of forming the MMA–amavadin adduct by a V–S bond. However, structures characterized by coordination of the SH group to vanadium do not correspond to energy minima. To explore further the possibility that S–V bond formation can take place, another scenario has been explored, in which the source of proton might be the –SH group itself.

This would imply a more favorable (S^-)–V bond formation. Indeed, even though the pH conditions (close to neutrality) at which Guedes da Silva *et al.* conducted their experiments should not allow the formation of deprotonated thiolate groups, it is not fair to exclude *a priori* that one of the O atoms of amavadin could act as a base. Another possibility might entail the transient formation of a deprotonated sulfur-protonated amavadin (transition) state with concerted (S^-)–V bond formation.

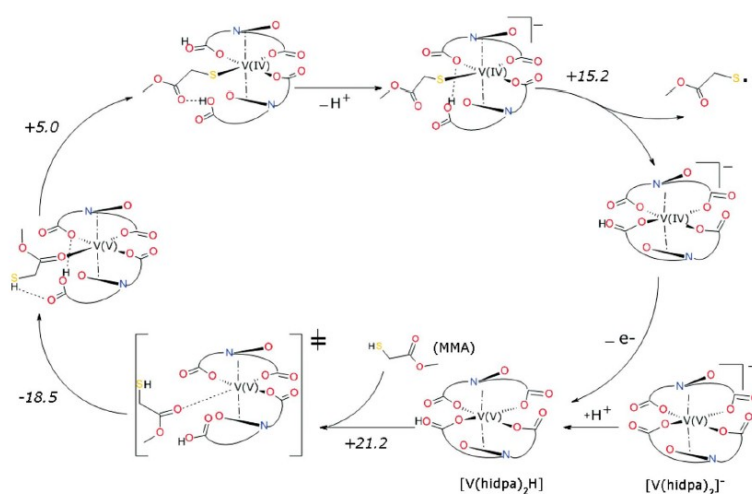
The most stable $[V(\text{hidpa})_2 \cdot \text{MMA}]^-$ adduct is characterized by proton transfer from the –SH group of MMA to one of the carboxylate groups of hidpa, and concomitant replacement of the carboxylic group in the vanadium first coordination sphere by the deprotonated S atom of MMA. However, it must be noted that the $\text{MMA} + [V(\text{hidpa})_2]^- \rightarrow [V(\text{hidpa})_2 \cdot \text{MMA}]^-$ reaction is endoergonic by more than 9 kcal mol⁻¹. Other $[V(\text{hidpa})_2 \cdot \text{MMA}]^-$ isomers featuring MMA coordination *via* the thiolate group lies lightly higher in energy (8.2–11.4 kcal mol⁻¹), essentially due to different intramolecular hydrogen bonding patterns. Notably, an isomer featuring simultaneous coordination of the thiolate and ester groups of MMA, replacing a carboxylic and a carboxylate ligand of $(\text{hidpa})_2$, is 27.1 kcal mol⁻¹ higher in energy relative to the most stable form. The V–S bond distance is 2.413 Å, which is

on average 0.4\AA longer with respect to V–O, but the $\delta(\text{V–S})$ index is 0.811 electron pair, which is roughly 0.1–0.2 electron pairs higher than average $\delta(\text{V–O})$ values computed for oxygen atoms that belong to carboxylate group. To discuss the atomic net charges of this species, we need to consider again the V and S charges in amavadin and MMA. To estimate the redox state of VN, we used a simple linear model ($\Omega(\text{V}) = 0.12 \cdot N + 1.402$) obtained by interpolation of the $\Omega(\text{V})$ values in V^{IV} and V^{V} amavadin. The V net charge in the anionic adduct is 1.901, which is less positive than that in V^{V} amavadin by 0.102 electron and corresponds to the $\text{V}^{4.3}$ redox state according to the linear model. To properly discuss this result, we computed the QTAIM charges for the anionic form of the MMA. The $\Omega(\text{S})$ value of -0.785 electrons testifies to the localization of the negative charge in MMA- on the sulfur atom, as one could expect. When MMA is coordinated through sulfur to the V^{V} of amavadin, the S atom is less negatively charged by 0.469 electron, with a $\Omega(\text{S})$ value of -0.316. The two results presented indicate that upon MMA coordination the coordinated S atom is partially oxidized and conversely V is partially reduced; a fraction of the S electron density is shifted toward the V–S region to form the V–S bond. The remaining sulfur electron density is essentially delocalized on the oxygen atom of the V coordination, as shown by the average increase of their negative charges. Finally, it is interesting to note that according to the model (4), the VV redox state in the most stable neutral intermediate does not change

4.2.3.5. *Final mechanistic considerations on thiol oxidation*

The DFT characterization of reaction intermediates formed when MMA reacts with amavadin allowed us also to shed some light on the catalytic

mechanism (Scheme 3). According to our results, only the protonated V^V form of amavadin $[V(\text{hidpa})_2\text{H}]$ can bind MMA, forming an intermediate species in which the carbonyl MMA group acts as a vanadium ligand. The latter observation is in agreement with the experimental observations indicating the necessity of a carbonyl moiety for turnover. The lowest energy $\text{MMA}-[V(\text{hidpa})_2\text{H}]$ isomer can subsequently convert to an isomer featuring direct coordination of the S atom of MMA to vanadium. Notably, the formation of the latter intermediate species corresponds to the redox step of the mechanism, in which the oxidation state of vanadium changes from V to IV and the MMA molecule is oxidized. The catalytic cycle for monoelectron oxidation of MMA is closed by release of the thiol radical and oxidation of the catalysts. Radical coupling of two oxidized MMA molecule leads to S–S bond formation.



Scheme 3 Proposed catalytic cycle for the reaction: $\text{MMA} \rightarrow \text{MMA}^\bullet + \text{H}^+ + \text{e}^-$ as turnover by the protonated form of amavadin. Energies are indicated in kcal mol^{-1} . Those steps corresponding to protonation/deprotonation reactions and oxidation reactions are not labelled with energetic tags since, by presently available DFT theory levels, both $\text{p}K_a$ and reduction potential calculations are affected with too large margins of error to be directly compared to experimentally observed values.

4.2.4. Conclusions

In this part of the thesis light was shed on energetic and structural features of an adduct which forms between amavadin and a prototypical substrate (MMA) in the thiol oxidation to the corresponding disulfide, a reaction of extreme importance in biological systems. In particular, DFT shows that both thermodynamics (slightly endoergonic reaction) and kinetics (non-prohibitive reaction barrier) of the adduct formation are in line with an accessible process. QTAIM and structural parameters are consistent with the activation of the substrate when bound to the catalyst. Moreover, an indirect evidence on the role of the vanadium (in its d^0 active state) based catalyst protonation has been proposed.

Interestingly, theoretical results have allowed us to rationalize, on the basis of structure–function relationships, one of the most relevant experimental observations by Guedes da Silva *et al.*: an ancillary carboxylate group (or its derivative ester function) is a crucial requisite in order to detect electrocatalytic activity. Thus, the identification by DFT of a species in which V^V is ligated by one of the ester oxygen atoms of MMA is strongly supportive of the role played in the whole oxidative process by this secondary functionality present on reactive substrates. Secondly, experimental data pointed out a precise geometrical requirement which had to be fulfilled by active substrates: the ester/acid moiety must be two or three bonds away from the reducing–SH center. In regard to this aspect, DFT results permit only speculative conclusions to be drawn. Actually, an *unstable* isomer adduct has been located (Scheme 1) which features a bis-chelate [O;S] coordination fashion of MMA to V, which implies a favorable intramolecular 5-membered ring formation. If methyl mercaptopropionate (another reactive substrate) were to replace MMA, one should conceivably

assist a 6-membered ring formation. According to a Michaelis–Menten catalysis scheme, after the substrate–catalyst adduct has formed, its destabilization must occur to observe reactions. Such unstable forms with chelate coordination of MMA might represent such reactive species. Finally, some stereoelectronic features of the catalyst itself, already investigated in the past with less-sophisticated methodologies, have been revisited (and confirmed) by means of high-level theory calculations.

This work lead to the following publication:

DFT characterization of key intermediates in thiols oxidation catalyzed by amavadin, *Dalton Transaction*, 40 (30): 7704-7712, 2011

4.2.5. References

- 1) J. H. Trefy and S. Metz, *Nature*, 1989, **342**, 531.
- 2) D. Rehder, *Org. Biomol. Chem.*, 2008, **6**, 957.
- 3) E. Bayer and H. Kneifel, *Z. Naturforsch B*, 1972, 207.
- 4) H.-U. Meisch *et al.*, *Naturwissenschaften*, 1979, **66**, 620.
- 5) M. A. A. F. de C. T. Carrondo *et al.*, *J. Chem. Soc., Chem. Commun.*, 1988, 1158.
- 6) E. M. Armstrong *et al.*, *J. Am. Chem.Soc.*, 1993, **115**, 807.
- 7) H. Kneifel and E. Bayer, *Angew. Chem., Int. Ed. Engl.*, 1973, **12**, 508.
- 8) C.M.M.Matoso *et al.*, *ACS Symp. Ser.*, 1998, **711**, 241.
- 9) P. M. Reis, *et al.*, *Chem. Commun.*, 2000, 1845.
- 10) M. V. Kirillova *et al.* *J. Am. Chem. Soc.*, 2007, **129**,10531.
- 11) M. V. Kirillova *et al.*, *Chem.–Eur. J.*, 2008, **14**, 1828–1842.
- 12) T. Hubretsge *et al.*, *Eur. J. Org. Chem.*, 2007, 2413–2422.
- 13) R. D. Thackerey and T. L. Riechel, *J. Electroanal. Chem.*, 1988, **245**, 131.
- 14) M. F. C. G. da Silva *et al.*, *J. Am. Chem. Soc.*, 1996, **118**, 7568.

- 15) E. M. Armstrong *et al.*, *J. Chem. Soc., Dalton Trans.*, 1995, 191.
- 16) C. Remenyi *et al.*, *J. Phys. Chem. B*, 2005, **109**, 4227.
- 17) K. J. Ooms *et al.*, *Dalton Trans.*, 2009, (17), 3262–3269.
- 18) K. R. Geethalakshmi *et al.*, *Inorg. Chem.*, 2007, **46**, 11297.
- 19) D. C. Crans *et al.*, *Inorg. Chem.*, 2010, **49**, 4245.
- 20) A. D. Becke, *Phys. Rev. A: At., Mol., Opt. Phys.*, 1988, **38**, 3098–3100; J. P. Perdew, *Phys. Rev. B.*, 1986, **33**, 8822.
- 21) K. Eichkorn *et al.*, *Theor. Chem. Acc.*, 1997, **97**, 119.
- 22) R. Ahlrichs *et al.*, *Chem. Phys. Lett.*, 1989, **62**, 165.
- 23) A. Schafer *et al.*, *J. Chem. Phys.*, 1994, **100**, 5829.
- 24) G. Zampella *et al.*, *Inorg. Chem.*, 2006, **45**, 7133.
- 25) G. Zampella, *et al.*, *J. Am. Chem. Soc.*, 2005, **127**, 953.
- 26) A. Klamt, *J. Phys. Chem. A*, 1995, **99**, 2224.
- 27) A. Klamt, *J. Phys. Chem. A*, 1996, **100**, 3349.
- 28) A. Klamt and G. Schüürmann, *J. Chem. Soc. Perkin Trans.*, 1993, **2**, 799.
- 29) F.W. Beigler-Konig *et al.*, *J. Comput. Chem.*, 1982, **317**, 3.
- 30) C. Gatti, SF_ESI Code, private communication.
- 31) X. Fradera *et al.*, *J. Phys. Chem. A.*, 1999, **103**, 304.
- 32) P. Macchi and A. Sironi, *Coord. Chem. Rev.*, 2003, **238–239**, 383.
- 33) C. Gatti and D. Lasi, *Faraday Discuss.*, 2007, **135**, 55.
- 34) L. Bertini *et al.*, *Organometallics*, 2010, **29**, 2013–2025.
- 35) R. E. Berry, *Angew. Chem. Int. Ed.*, 1999, **38**, 795.
- 36) C. Lee *et al.*, *Phys. Rev. B*, 1988, **37**, 785.
- 37) S. Vosko *et al.*, *Can. J. Phys.*, 1980, **58**, 1200.
- 38) A critical point is a stationary point of the electron density characterized by the rank (the sum of the non-zero hessian eigenvalues) and the signature (the sum of the sign of the element of the diagonalized hessian). A bond critical point (BCP) denotes a (3,-1) first-order saddle point of electron density along the maximum electron density line between two atoms.

4.3. DFT investigation of the copper coordination to the putative cell binding site of angiogenin

4.3.1. The biological question

Angiogenesis is a set of functional processes responsible for the formation of new blood vessels from existing ones, and is also a key component of the homeostatic process that regulates the distribution of oxygen to tissues [1]. Angiogenesis occurs in several tightly regulated stages that orchestrate a network of cooperative interactions, and can be divided in several phases: [1] an initiation phase, characterized by increased vasopermeability; [2] a progression phase, in which proteolytic enzymes that degrade the extracellular matrix and promote endothelial cell migration are produced; [3] a final phase, in which differentiation into new vessels takes place. The latter phase is mediated by molecules that recruit mesenchymal cells to vessel walls [2].

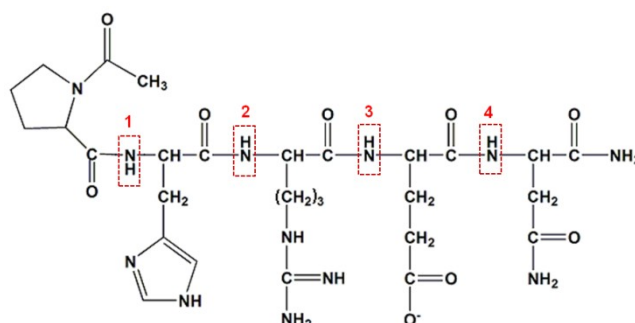
Angiogenin (Ang) is an angiogenic protein that undergoes nuclear translocation in endothelial cells where it accumulates in the nucleolus and stimulates rRNA transcription, a rate limiting step in ribosome biogenesis, protein translation, and cell growth [3]. Ang was originally isolated from the conditioned medium of HT-29 human colon adenocarcinoma cells [4]. The mature Ang is a basic, single-chain protein containing 123 amino acids with a molecular weight of about 14.400 Da, and it is a homologue of bovine pancreatic ribonuclease A. Indeed, its ribonucleolytic activity is rather low, but Ang is essential for angiogenesis and other functions [5]. In fact, Ang belongs to the ribonuclease superfamily, showing 33% sequence identity to

the pancreatic ribonuclease A. However, although the crystal structures of human Ang and pancreatic ribonuclease A are highly similar, a notable difference is evident in the ribonucleolytic active site: the pyrimidine binding site of Ang is “obstructed” by the Gln117 residue, explaining its very weak ribonucleolytic activity (about 10^5 – 10^6 lower than that of RNase A).

Besides its ribonucleolytic activity, binding of Ang to the endothelial cells surface is needed for its biological functions, and amino acid residues from 60 to 68 are critical in this process. Notably, affinity of Ang for endothelial cells is largely increased in presence of copper ions [6]. In fact, copper has been recognized to be an angiogenic factor, but the mechanism whereby it exerts this function is still not well understood [7, 8]. In particular, Ang binds 2.4 mol of copper per mol of protein, and the metal ion is important for Ang binding to calf pulmonary artery endothelial cells, which increases by 4.3 fold in the presence of Cu^{2+} .

A putative endothelial cell binding domain of Ang is located in a loop exposed to the solvent and largely unstructured in the native protein [9]. The related amino acid sequence (hAng60–68 = KNGNPHREN) contains the prolyl-histidyl (PH) dyad which is reminiscent of that present in other copper binding proteins, such as prion proteins [10] and the Wilson’s and Menkes’ ATPases [11]. Indeed, the comparison between Ang and RNase A structures already suggested possible copper binding sites [12]. In fact, RNase A binds copper at several sites including His-12, His-105 and His-119 at pH 5 [13], but Ang features three additional histidines which are not present in RNase A: His-8, His-65 and His-84 [14, 15]. The similarity between the copper binding plasma tripeptide Gly-His-Lys and the copper binding site of albumin and α -fetoprotein, where copper binds to a histidyl residue adjacent to a basic residue (Arg or Lys), suggests that His-65, which is adjacent to Arg-66 in Ang, might be involved in copper binding.

Very recently Bonomo and collaborators [16] reported the synthesis and the physico-chemical characterization of the complex formed by the peptide fragments encompassing the sequence hAng64–68 (Ac-PHREN-NH₂; Scheme 1) and the whole sequence hAng60–68 (Ac-KNGNPHREN-NH₂) with copper.



Scheme 1. Structure of the Ac-PHREN-NH₂ peptide. The four NH group of the peptide bonds have been highlighted in red boxes.

In particular, combined potentiometric and spectroscopic investigations allowed to reveal the species distribution and the coordination environments of the copper(II) complexes. It turned out that both peptides coordinate Cu²⁺ in a similar fashion. Moreover, thermodynamic and spectroscopic data indicated that the side chains of Glu and His residues are involved in copper binding at physiological pH. The copper(II) interaction with the peptide fragment Ang64–68(E67Q) (Ac-PHRQN-NH₂), in which glutamate was substituted by a glutamine residue, was also studied in order to unveil the role of glutamate carboxylate group on copper(II) coordination. The comparison between results obtained studying the copper(II) complexes formed by Ac-PHREN-NH₂ and its E67Q variant provided further evidence

of the presence of a carboxylate oxygen atom in the copper coordination sphere. On the ground of such results it was concluded that at pH 8 the metal coordination environment in the Ac-PHREN-NH₂/Cu²⁺ complex is formed by a nitrogen atom of His, two deprotonated amide groups, a water molecule and an oxygen atom from the COO⁻ side chain of Glu. However, it was not possible to reveal whether the carboxylate oxygen atom was coordinated in equatorial or apical position, hindering the full disclosure of the effects of Cu²⁺ binding on the structural properties of the Ang60-68 protein portion.

With the aim of complementing available experimental data in the elucidation of the nature of the copper first coordination environment, as well as the apical/equatorial disposition of oxygen ligands, we have carried out a DFT investigation of the copper(II) complexes formed by Ac-PHREN-NH₂, as well as by the single point mutated peptide Ac-PHRQN-NH₂.

4.3.2 Computational methods

Computations were performed using the pure Generalized Gradient Approximation (GGA) BP86 DFT functional [17, 18] and the Resolution of Identity (RI) technique [19], as implemented in the TURBOMOLE suite of programs [20]. Basis sets of triple-zeta plus polarization split valence quality (TZVP hereafter) [21] were adopted for all atoms in the complexes. The DFT grid-size was set to standard m3 value.

Molecular dynamic simulations were performed at 100K using the MOE suite (MOE Molecular Operating Environment, version 2008.10; Chemical Computing Group Inc.: Montreal, Canada, 2008) and using the MMFF94x forcefield [22, 23]. Each MD simulation were carried out in vacuum for 1000 ps.

The computation of the EPR g tensors and hyperfine coupling constants

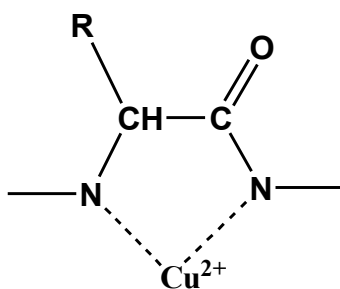
(hcc) for the ^{63}Cu nucleus have been performed using the B3LYP hybrid functional on the geometry optimized at the RI-BP86/TZVP level. An extended basis set (14s,10p,5d) [24], augmented by a set of diffuse s, p and d functions (with exponents equal to 0.01, 0.03087, and 0.1, respectively) and contracted to (9s, 7p, 4d) was adopted for the Cu atom. The IGLO-II basis was adopted for all other atoms [25]. EPR hyperfine coupling constants (hcc) were calculated, explicitly taking Spin-Orbit (SO) contributions into account, with the one-centre and mean field approximation (AMFI) [26] for the two electron terms [27]. For the calculations of g tensors the gauge origin has been set to the centre of electronic charge. Calculations of EPR properties have been carried out using the ORCA suite of programs [28].

4.3.3. Results and Discussion

4.3.3.1 Geometries and relative energies of isomers of Ac-PHREN-NH₂/Cu²⁺·H₂O and Ac-PHRQN-NH₂/Cu²⁺·H₂O complexes.

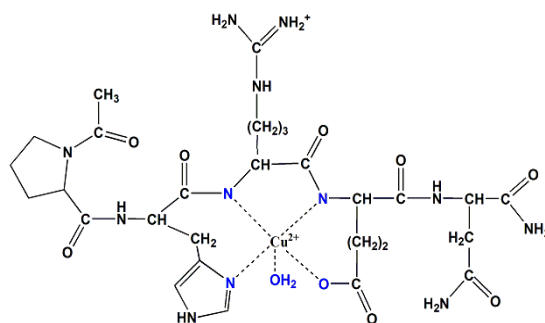
According to the experimental evidences, Cu²⁺ coordination to the Ac-PHREN-NH₂ peptide at pH 8 involves two N⁻ anions from the deprotonate amide groups of the peptide bonds, the carboxylate group of Glu and a nitrogen atom from the His side chain. In addition, a water molecule should also be coordinated to the metal ion [16]. Before discussing which deprotonated N⁻ groups are actually involved in Cu²⁺ coordination, some considerations are in order. Previous results obtained studying coordination compounds formed by Cu²⁺ with fragment of the prion protein [29-33] as well as simple considerations about the preferential formation of 5- and 6-member rings when dealing with chelate ligands, lead to the conclusion that

the two N^- anions involved in Cu coordination must be contiguous. Therefore, only three configurations are possible, namely those involving the **1-2**, **2-3** or **3-4** deprotonated amide groups (see Scheme 1 for amide groups labeling). We preliminary analyzed the conformation of these three configurations at Molecular Mechanics (MM) level, in order to quickly understand which is the most likely Cu^{2+} coordination geometry. The preliminary MM analysis revealed that, when the **1-2** or **3-4** configurations are taken into account, the histidine (in the **1-2** configuration) or the glutamic (in the **3-4** configuration) side chains must necessarily occupy the apical position in the metal coordination sphere. In fact, since the side chains of His (in the **1-2** configuration) or Glu (in the **3-4** configuration) are in between the deprotonated N^- groups that form the planar 5-member cycle (Scheme 2), the aminoacid side chain must necessarily coordinate Cu^{2+} in apical position due to steric restrictions. As a consequence, in complexes characterized by **1-2** or **3-4** binding mode the water molecule must necessarily occupy an equatorial position in the Cu^{2+} coordination environment.



Scheme 2. Five-member ring formed in Ac-PHREN-NH₂ upon coordination to Cu^{2+} of the two adjacent deprotonated amide groups.

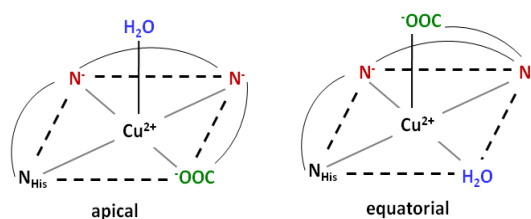
MM results indicate also that the **1-2** and **3-4** binding modes induce some strain in the peptide chain. In particular, the side chain of the coordinating aminoacid which does not belong to the five member ring (Glu in **1-2** complexes, His in **3-4** structures) are far from the metal center and their coordination to the metal atom is always accompanied by some strain of the peptide backbone. MM analysis of the **2-3** binding mode reveals that the imidazole ring of His and the carboxylate group of Glu can occupy equatorial positions in the Cu^{2+} coordination sphere without inducing steric strain in the molecule. These results suggest that the structures featuring the **2-3** binding mode most likely corresponds to the most stable isomers (see Scheme 3).



Scheme 3. Schematic representation of the **2-3** binding mode in complexes between Cu^{2+} , the peptide Ac-PHREN- NH_2 and a water molecule.

Therefore, we have carried out a thorough DFT investigation of the conformational properties of coordination compounds featuring the **2-3** binding mode. However, for the sake of completeness, we have evaluated by DFT also some structures characterized by the **1-2** or **3-4** binding mode, to quantify the strain of the peptide chain.

To select reasonable starting structures for DFT optimization, we have carried out classical MD simulations of the Ac-PHREN-NH₂/Cu²⁺·H₂O complex featuring the **2-3** binding mode, selecting snapshots along the MD trajectory which correspond to low energy conformations. The protocol adopted to explore the configurational space of the system is the following. The structures characterized by the **2-3** binding mode can be ideally divided into two families: the first containing structures featuring an apical water molecule (and consequently an equatorial Glu side chain), and the second containing structures featuring an equatorial water molecule (Scheme 4). Note that apical coordination of the His side chain is sterically impossible.



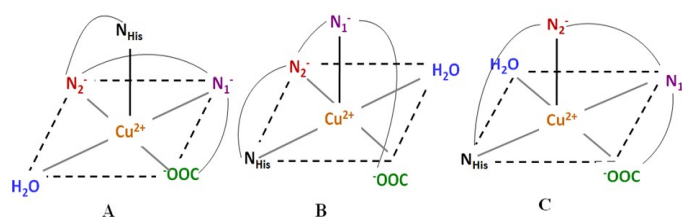
Scheme 4. The two families of Ac-PHREN-NH₂/Cu²⁺·H₂O complexes characterized by the **2-3** binding mode. Apical and equatorial labels are referred to the position of the water molecule in the Cu²⁺ coordination sphere.

We initially generated “ideal” (dihedral angle of the base of the pyramid equal to zero, molecular axes of the apical ligand exactly perpendicular to the base) structures characterized by the N₃O₂ coordination environment and featuring an apical or equatorial water molecule. The value of Cu²⁺-N⁻ distances has been set to 2.000 Å, according to the BP86 optimized Cu-N distances in HGGG/Cu²⁺ system [33]. For the Cu²⁺-N_{His} and Cu²⁺-O⁻_{Glu} distances, we adopted the value of 1.894 and 2.007 Å obtained respectively from BP86/TZVP gas-phase optimization of the Cu²⁺-His and Cu²⁺-Glu

adducts. Regarding the Cu-O distance to describe Cu-H₂O binding, we considered as experimental reference the XRD value of 2.4 Å in cis-Cu(Gly)₂·H₂O, which fits well also with the analysis of Sobolovic et al. [34], which measured a mean value of 2.4±0.1 Å for Cu-O distances in 10 aminoacid/water copper(II) complex XRD structures.

Using the two “ideal” structures as templates, we generated apical and equatorial conformations from a molecular dynamics simulation in which the Cu²⁺ coordination distances were constrained to the above mentioned reference values. Then, for each structure obtained, we first optimized its geometry at the MM level of theory, and the corresponding minimum energy structure was used as starting point of a two-step DFT optimization. In the first step the Cu²⁺ coordination was constrained, allowing only the peptide moiety to relax. In the second step the whole geometry was optimized removing the constraints.

The relevant optimized geometry parameters, as well as computed energies of all structures characterized by the 2-3 binding mode are collected in Table 1. It must be noted that in structures characterized by an apical water molecule the N₃O₂ coordination can be realized only in one way, whereas in the case of equatorial water binding three different possibilities, referred to as the A, B and C subfamilies hereafter, can in principle exist (Scheme 5).



Scheme 5. The three subfamilies of Ac-PHREN-NH₂/Cu²⁺·H₂O complexes characterized by the 2-3 binding mode and the water molecule in equatorial position.

Before discussing the fine details of the various structures optimized, it is interesting to highlight some trends obtained by our DFT investigation. Except for a few structures, geometry optimizations usually led to tetrahedral distortion of the initial square pyramidal coordination. In addition, the H₂O ligand often resulted loosely bound or non-bound to Cu²⁺, yielding 4-coordinated N₃O structures. It is also worth noting that when starting from structures featuring an equatorial water ligand, the H₂O molecule always leaved the Cu²⁺ coordination sphere, and it was replaced in the equatorial position by the oxygen atom of the glutammic acid residue. Therefore, it can be concluded that structures belonging to the subfamilies A, B or C do not correspond to stable forms of the Ac-PHREN-NH₂/Cu²⁺ complex.

In the exploration of the conformation space of Ac-PHREN-NH₂/Cu²⁺, we often found that structures can differ only for the number of intramolecular H-bonds. As shown in the following, when a complex is explicitly solvated by a reasonable number of H₂O molecules, intramolecular interactions are often lost because the side-chains of the residues not involved in Cu²⁺ coordination (in particular arginine) prefer to form H-bonds with the solvent. Nevertheless, even if we are aware of the limits of such approach, we carried out most DFT calculations neglecting solvent effects. In fact, the explicit description of a shell of water molecules surrounding the Ac-PHREN-NH₂/Cu²⁺ complex would be computationally too demanding. Indeed, since one the main purposes of this study was the elucidation of the position of the H₂O ligand in the Cu²⁺ coordination sphere, in principle solvent effects due to bulk water could be described using an implicit solvent model. However, it was already observed that DFT tends to underestimate H₂O-Cu²⁺ binding in similar complexes [33], and we noted that the adoption of an implicit solvent model makes the H₂O-Cu²⁺ bond even weaker, not allowing to properly study H₂O coordination to Cu²⁺ in this class of

compounds.

The lowest energy structure characterized by the **2-3** binding mode features an apical water molecule (**1**; Figure 1). In **1** the N_3O_2 square pyramidal Cu^{2+} coordination environment is characterized by a dihedral angle $\hat{\curvearrowright}$ (defined by the atoms $N_{HIS}-N^--N^--O^-$ forming the base of the square pyramid) equal to 0.4 degree, and a $Cu-O_{H_2O}$ distance = 2.645 Å. Other five structures are characterized by a $Cu-O_{H_2O}$ distance lower than 3.0 Å (**8**, **11**, **12**, **14**, **19**; Table 1). When compared to **1**, these structures are characterized by a smaller number of H-bonds, except for structure **14**, which however is characterized by a loosely bound H_2O molecule ($Cu-O_{H_2O} > 2.753 \text{Å}$). The structure with the shortest $Cu-O_{H_2O}$ distance (**12**) features an apical water molecule and is 13.7 $\text{kcal}\cdot\text{mol}^{-1}$ higher in energy than **1**. In this context it is also worth noting that **1** and **12** differ for the number of intramolecular H-bonds (4 in **12**, 6 in **1**; $d(H\cdots O) \leq 2.0 \text{Å}$), underlying the role played by this type of interactions for the relative energies of the structures considered.

The electronic spectra of Ac-PHREN-NH₂ and the lowest energy Ac-PHREN-NH₂/Cu²⁺·H₂O form (structure **1**) have been computed at BP86/TZVP TDDFT level and are shown in Figure 2. The first 100 transitions have been computed and excitation energies, corresponding oscillation strengths and MO compositions for the first 30 transitions were collected. It must be noted that the BP86 functional is not the optimal choice to reproduce excitation energies, in particular when metal-to-ligand charge transfer (MLCT) bands are considered, but the large size of the systems investigated did not allow us to adopt more suited functional, as PBE0 [35]. In any case, computed UV-Vis bands of Ac-PHREN-NH₂ are mainly due to histidine residue, and correspond to two intense absorptions at 266 nm and 240 nm, and one weak band at 320 nm. In the corresponding Cu²⁺ complex the two intense bands are red-shifted by about 100 nm. The bands in the

range 1000-500 nm (Table 2) are due to Cu^{2+} and correspond to ligand-to-metal charge transfer bands. In this range the most intense band is at 689 nm, nicely fitting with the 705 nm absorption reported by La Mendola et al. [16], and corresponds to a charge transfer transition from to the oxygen atom of the Glu side chain to the copper ion.

As already pointed out, structures characterized by equatorial disposition of the water molecule have not been identified, suggesting an intrinsic instability of this coordination mode. In order to quantify the destabilization of structures featuring an equatorially coordinated water molecule, we have also optimized a structure (**20**) which is obtained from **1** after exchange of the positions of the water molecule and the carboxylate group. In order to maintain the Cu^{2+} coordination environment, the position of the two coordinated oxygen atoms from the water molecule and the carboxylate group have been constrained to the Cu-O distances observed in **1**. Although this evaluation is qualitative, we found that structure **20** is more than 30 $\text{kcal}\cdot\text{mol}^{-1}$ higher in energy than **1**.

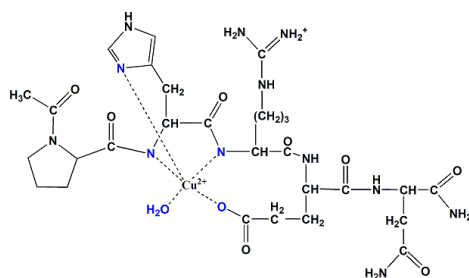
In order to evaluate if an extra water molecule might be bound to Cu^{2+} , leading to a square bipyramidal coordination environment, we further investigated complex **1** after addition of a second H_2O molecule in apical position. However, the optimization of such a species led to a square pyramidal structure in which the extra water molecule moved away from the metal ion (final Cu-O_w distance = 3.567Å).

Going back to the issue of the computational approach adopted in this study to describe solvation effects, we also optimized the most stable structure (**1**) including solvent effects through the implicit solvent model COSMO (COnductor-like Screening MOdel) [36-38]. The main effect observed in COSMO calculations has been the removal of the H_2O molecule from the Cu^{2+} coordination sphere, which consequently becomes tetrahedrally

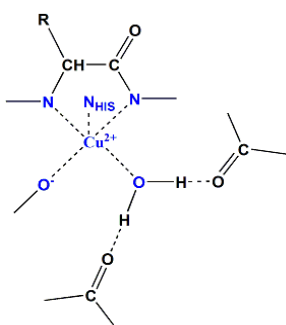
distorted (δ moves from 0.4 to 30.7 degree), in disagreement with experimental results [16], and confirming the validity of the gas-phase approach adopted to study Ac-PHREN-NH₂/Cu²⁺ complexes.

Finally, we have also evaluated the lowest energy complex **1** when explicitly solvated. To address this point, structure **1** was solvated with a cluster of 55 H₂O molecules, whose position was optimized by mean of a short MD simulation (1500 ps) and successive MM geometry optimization. The solvated structure obtained in this way was successively optimized at DFT level using the same level of theory adopted for the non-solvated models (BP86/TZVP). Upon geometry optimization, the Cu²⁺ coordination does not change significantly. The H₂O molecule initially at 2.645Å gets closer to the metal center (2.445Å) and the δ dihedral angle increases of only 2 degree. Notably, no intramolecular H-bonds are found because the aminoacid sidechains are all involved in H-bond with the solvent.

DFT results obtained studying the **1-2** and **3-4** forms indicate that these isomers are higher in energy with respect to the most stable **2-3** isomer (**1**). In particular, most of the **1-2** structures (**21**, **22**, **23**, **26** and **28**; Table 1) feature a Cu²⁺ coordination with the histidine residue in apical position (see Scheme 6) and are higher in energy with respect to isomer **1** by at least 13.7 kcal·mol⁻¹.



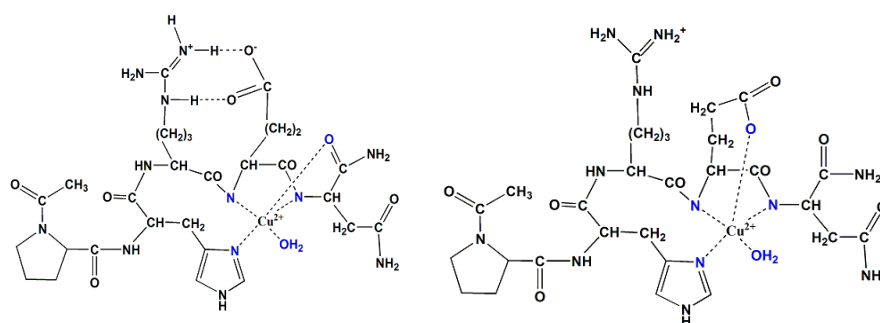
The three lowest energy **1-2** structures feature square pyramid coordination with a small tetrahedral distortion. The lowest energy **1-2** form (structure **21**) is shown in Figure 3. It is interesting to point out that the H₂O ligand is coordinated to Cu²⁺ with a much shorter Cu-O distance (2.031Å) compared with the **2-3** structures. In fact, in **21** the Cu²⁺-H₂O binding is “assisted” by the formation of two H-bonds formed with the two CO groups of the Glu sidechain and the peptide bond between Pro and His (Scheme 7).



Scheme 7. Schematic representation of the Cu²⁺ coordination environment in the lowest energy complex characterized by the **1-2** binding mode (structure **21**).

During the conformational sampling of structures characterized by the **1-2** binding mode we have found three 4-coordinated structures (**24**, **25** and **27**), in which the histidine sidechain is not coordinated to Cu²⁺. However, these isomers are at least 8.6 kcal·mol⁻¹ higher in energy than the lowest energy **1-2** form (**21**).

The analysis of the MD simulations of the **3-4** configuration revealed a larger number of possible copper coordinations. In particular, the structures obtained from the MD simulations can be classified in two families (Scheme 8).



Scheme 8. Schematic representation of the two families of complexes between Cu^{2+} , the peptide Ac-PHREN- NH_2 and a water molecule characterized by the **3-4** binding mode.

In structures **29**, **31**, **36** and **40** (Scheme 8; *left*) the carboxylate of the glutamic acid forms an electrostatic interaction with the side-chain of arginine, and consequently the carbonyl group of the C-terminal amide group enters the Cu^{2+} coordination sphere in equatorial position. Among the family *a* species, structure **29** is the lowest in energy ($+8.1 \text{ kcal}\cdot\text{mol}^{-1}$ with respect to **1**).

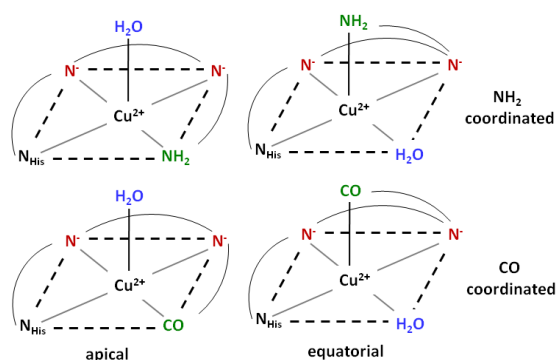
In structures **32**, **35**, **38**, **41** and **43** (Scheme 8; *right*) the carboxylate group of Glu is coordinated to Cu^{2+} in apical position, and therefore the water molecule occupy the equatorial position. All these structures are genuine equatorial forms, but, however, are higher in energy with respect to the constrained **2-3** equatorial form (structure **20**).

The other **3-4** structures obtained from the conformational sampling are higher in energy and feature a different Cu^{2+} coordination environment. It is worth noting that the 3-coordinated structure **30** (2N^- and N_{His}) lies only $1.6 \text{ kcal}\cdot\text{mol}^{-1}$ higher in energy with respect to the lowest energy **3-4** structure (**29**) because it is stabilized by the formation of 6 intramolecular H-bonds. Structures **33** and **34**, which are characterized by substitution of the

coordinated H₂O with a CO group of the peptide chain, lie at least 29.6 kcal·mol⁻¹ higher in energy with respect to **1**. In **34** also the histidine sidechain leaves the Cu²⁺ coordination sphere, giving a 4-coordinated form. Structure **37** is the only **3-4** form that features the histidine sidechain in apical position, similarly to the **1-2** forms.

In order to better understand the role of the glutamate side chain in Cu²⁺ binding, and prompted by the previous observation that cleavage of the E67-N68 peptide bond compromises the angiogenic activity of Ang [39], Bonomo and collaborators investigated also Cu²⁺ binding to the E→Q variant of the Ac-PHREN-NH₂ peptide, concluding that at pH 8 also in Ac-PHRQN-NH₂ the Cu²⁺ coordination is of N₃O₂ type, with the C=O group of glutamine providing the fifth ligand [16].

In principle, four different types of coordination can be observed in complexes between Cu²⁺ and Ac-PHRQN-NH₂ (Scheme 9), in which the glutamine residue can be coordinate to Cu²⁺ via the NH₂ or CO group of its side-chain, and the H₂O ligand can occupy apical or equatorial position.



Scheme 9. The four conformation subfamilies of the Ac-PHRQN-NH₂/Cu²⁺·H₂O complex. The labels apical and equatorial are referred to the position of the water molecule in the Cu²⁺ coordination sphere.

Sampling of the configuration space of the Ac-PHRQN-NH₂/Cu²⁺ complex has been carried out using the same approach adopted for the wild-type peptide. In particular, starting from the lowest energy forms of the Ac-PHREN-NH₂ complex, we introduced the E→Q substitution and successively carried out MD simulations to obtain starting point structures for DFT optimizations. We extracted from the MD trajectories 8 isomers characterized by apical (4 structures) or equatorial (4 structures) H₂O coordination, and for each structure we evaluated both CO and NH₂ coordination of the glutamine side-chain, obtaining a total number of 16 starting point structures for DFT optimization.

In general, in structures characterized by coordination of the NH₂ group of Gln, the Cu-N distance is 2.299 Å and the NH₂ group is in apical position. The corresponding form in which the CO amidic group is coordinated to Cu²⁺ features a Cu-O distance of 2.180 Å. Indeed, it is well known that the oxygen atom of an amide group is a better ligand than the NH₂ group. Accordingly, DFT models featuring coordination of the nitrogen amide atom are at least 15 kcal·mol⁻¹ higher in energy.

Relative energies of the various DFT optimized structures, as well as relevant bond distances for Ac-PHRQN-NH₂/Cu²⁺ complexes are collected in Table 3. Notably, all Ac-PHRQN-NH₂/Cu²⁺ low energy structures are characterized by a very evident tetrahedral distortion. The lowest energy Ac-PHRQN-NH₂/Cu²⁺ isomer (**1Q**, see Figure 4) features the H₂O molecule in equatorial position and shows strong tetrahedral distortion ($\delta=25.2$ degree). Moreover, in **1Q** the glutamine sidechain is coordinated to the Cu²⁺ via the carbonyl group, as expected. However, a four-coordinate isomer (**3Q**) in which the glutamine sidechain has left the Cu²⁺ coordination sphere is only 2.2 kcal·mol⁻¹ higher in energy. Similarly, the isomer **5Q** is also 4-

coordinated, but it is 10.3 kcal·mol⁻¹ higher in energy due to a smaller number of intramolecular H-bond interactions.

The structures **8Q** and **9Q** are similar and are the lowest in energy among those in which the NH₂ group of glutamine is coordinated to Cu²⁺. However their energy is 15.2 kcal·mol⁻¹ higher than in **1Q**. A number of higher energy 4-coordinated forms have been also found (structure **13Q-18Q** and **20Q**), in which the water molecule leaves the Cu²⁺ coordination sphere.

4.3.4. Conclusions

In this part of the thesis we contributed to the study of the Cu²⁺ coordination to the Ac-PHREN-NH₂ peptide, that represents the Cu²⁺ binding portion of angiogenin, a protein playing a key role in the process of angiogenesis. Using as a reference the recent experimental results reported by Bonomo and collaborators [16], we have used classical MM/MD calculations followed by DFT optimizations to explore the configurational space of the Ac-PHREN-NH₂/Cu²⁺·H₂O complex.

EPR experiments suggested that Cu²⁺ is coordinated in a N₃O₂ square pyramidal environment, where the Cu²⁺ coordination sphere is composed by 2N⁻ of deprotonated peptide bonds, the oxygen atom of a water molecule, the carboxylate group of the glutamic residue and a nitrogen atom of the histidine group. Experimental results did not allow to elucidate the relative position of water and carboxylate ligands within the Cu²⁺ coordination environment. However, a detailed comparison of the calculated minimum energy structures and EPR parameters suggests that i) binding of Cu²⁺ to the deprotonated amide groups of the peptide bonds takes place between the residues H-R and R-E, ii) the carboxylate group and the H₂O molecule are

coordinated in equatorial and apical position, respectively, in a slightly distorted square pyramidal arrangement.

The configuration space of the E→Q mutated system Ac-PHRQN-NH₂/Cu²⁺·H₂O has also been studied. In this case, computational results led to the conclusion that the H₂O molecule is coordinated in equatorial position and the oxygen atom of the carbonyl group of glutamine is weakly coordinated in apical position.

Table 1. Selected geometrical parameters (in Å and degree) and energy differences relative to the lowest energy form (in kcal·mol⁻¹) of the [Cu(Ac-PHREN-NH₂)(H₂O)] isomers. In the table are reported the five Cu-L distances, the dihedral angle that defines the base of the square pyramid of the N₃O₂ coordination sphere and the number of hydrogen bonds (H···O distance less than 2.0Å). CN stands for coordination number.

Isomer	CN	ΔE	Distances					∠ dihedral	H-bonds
			Cu-O _{H2O}	Cu-O _{CO}	Cu-N ₁	Cu-N ₂	Cu-N _{His}		
2-3 binding mode									
1	5	0.0	2.645	2.039	2.016	2.019	2.039	0.4	6
2	4	2.0	3.777	2.001	2.017	2.001	2.000	16.6	5
3	4	2.1	3.863	2.012	2.020	2.000	2.006	12.4	5
4	4	2.1	3.796	2.010	2.012	1.999	1.994	13.5	5
5	4	5.5	3.565	2.001	1.996	1.948	2.016	38.5	2
6	4	6.0	3.374	1.973	1.965	2.034	2.003	9.2	4
7	4	6.2	3.354	1.999	2.029	2.033	2.032	17.6	4
8	5	6.3	2.488	2.014	2.017	2.056	2.029	0.9	3
9	4	8.4	3.684	1.983	1.986	2.022	2.012	8.7	4
10	4	12.2	3.655	1.982	1.932	2.074	1.986	42.1	8
11	5	12.7	2.492	2.020	1.994	2.070	2.028	4.3	5
12	5	13.7	2.348	2.044	2.037	2.023	1.971	3.4	4
13	4	14.0	3.390	2.002	1.963	1.929	2.100	37.3	3
14	5	14.8	2.753	2.009	2.032	1.993	2.047	1.3	6
15	4	15.3	3.912	1.954	1.954	2.016	2.021	35.3	5
16	4	15.5	3.755	2.010	1.990	2.019	2.011	18.2	5
17	4	16.0	3.924	1.962	1.949	2.023	2.020	34.9	5
18	4	17.7	4.073	1.932	1.972	2.002	2.038	35.2	6
19	5	24.6	2.889	1.997	2.000	2.013	2.054	19.6	4

20	5	32.3	2.407	1.997	2.000	2.004	1.994	0.4	6
2-3 binding mode									
21	5	13.7	2.031	2.042	1.984	2.150	2.281	11.0	8
22	5	13.8	2.023	2.042	1.982	2.130	2.326	14.4	8
23	5	20.0	2.020	2.050	2.050	2.004	2.302	9.6	4
24	4	22.3	1.994	2.041	1.977	1.976		44.0	5
25	4	25.4	1.979	2.070	1.987	1.968		47.2	6
26	5	26.7	2.014	2.112	2.029	2.038	2.237	38.9	6
27	4	26.8	1.973	1.998	1.998	1.985		30.6	3
28	5	31.2	2.017	1.989	2.107	2.063	2.084	47.5	4
3-4 binding mode									
29	5	8.1	2.450	2.242*	1.931	1.999	2.050	31.0	5
30	3	9.7	3.075		1.915	2.023	1.978		6
31	5	16.4	2.544	2.178*	2.012	1.927	2.024	19.7	5
32	5	26.1	2.170	2.010	2.000	2.078	2.168	47.5	5
33	5	26.2	3.414	2.574 2.711*	1.935	1.986	2.003	53.7	6
34	5	35.2	3.641	2.085 1.996*	1.911	1.993		28.8	5
35	5	36.1	2.035	2.054	1.999	2.073	2.294	14.3	5
36	5	37.3	2.289	2.627*	2.071	1.923	2.018	16.3	5
37	5	38,1	2.142	2.088	2.003	2.054	2.227**	21.0	5
38	5	39.4	2.041	2.142	1.984	2.058	2.287	55.6	4
39	5	41.6	5.610	1.995	1.969	2.038	2.033	36.9	6
40	5	42.4	2.289	2.267*	2.020	1.940	2.035	16.3	3
41	5	50.6	2.165	2.306	2.000	2.023	2.025	20.6	3

42	5	51.2	3.889	1.993	1.983	2.053	2.222	58.3	3
----	---	------	-------	-------	-------	-------	-------	------	---

* The oxygen atom coordinated to Cu²⁺ does not belong to the COO⁻ group of the Glu side chain residue;

** Histidine in apical position;

This work lead to the following publication:

Copper coordination to the putative cell binding site of angiogenin. A DFT investigation., *Inorg.Chemistry. Accepted*

4.3.5. Acknowledgments

This work was supported by the PRIN Project N 200875WHMR.

4..3.6. References

- 1) Folkman J., *J Natl Cancer Inst*, 1989, **82**:4
- 2) Bussolino F. et al., *Trends Biochem Sci*, 1997, **22**:251
- 3) Kishimoto K. et al., *Oncogene*, 2005, **24**:445
- 4) Fett J.W. et al., *Biochemistry*, 1985, **24**:5486
- 5) Gao X. and Xu Z., *Acta Biochim Biophys Sin*, 2008, **619**
- 6) Badet J. et al, *Proc Natl Acad Sci USA*, 1989, **86**:8427
- 7) McAuslan B.R. and Reilly W., *Exp Cell Res*, 1980, **130**:147
- 8) Hu G-F., *J Cell Biochem*, 1998, **69**:326
- 9) Lequin O. et al., *Eur J Biochem*, 1997, **50**:712
- 10) Millhauser G.L., *Annu Rev Phys Chem*, 2007, **58**:299
- 11) Solioz M. and Vulpe C., *Trends Biochem Sci*, 1996, **21**:237
- 12) Soncin F. et al., *Biochem Biophys Res Commun*, 1997, **236**:604
- 13) Joyce B.K. and Cohn M., *J Biol Chem*, 1969, **244**:811
- 14) Acharya K.R. et al., *Proc Natl Acad Sci*, 1994, **91**:2915
- 15) Smyth D.G. et al., *J Biol Chem*, 1963, **238**:227

- 16) La Mendola D. *et al.*, *Dalton Trans*, 2010, **39**:10678
- 17) Becke A.D., *Phys Rev A*, 1988, **38**:3098
- 18) Perdew J.P., *Phys Rev B*, 1986, **33**:8822
- 19) Eichkorn K. *et al.*, *Theor Chem Acc*, 1997, **97**:119
- 20) Ahlrichs R *et al.*, *Chem Phys Lett*, 1989, **62**:165
- 21) Schafer A. *et al.*, *J Chem Phys*, 1994, **100**:5829
- 22) Halgren T.A., *J Comput Chem*, 1996, **17**:490
- 23) Halgren T.A., *J Comput Chem*, 1999, **20**:720
- 24) Schafer A. *et al.*, *J Chem Phys*, 1992, **97**:2571
- 25) Kutzelnigg W. *et al*, 1990, Springer-Verlag
- 26) Hess BA *et al.*, *Chem Phys Lett*, 1996, 251: 365
- 27) Kaupp M. *et al.*, 2004, Wiley-VCH
- 28) Neese F., *J Chem Phys*, 2003, **117**:3939
- 29) Marino T. *et al.*, *J Phys Chem B*, 2007, **111**:635
- 30) Marino T. *et al.*, *Int J Quantum Chem*, 2011, **111**:1152
- 31) Pushie M.J. and Rauk A., *J Biol Inorg Chem*, 2003, **8**:53
- 32) Franzini E. *et al.*, *Inorg Chem Commun*, 2003, **6**:650
- 33) Bruschi M. *et al*, *Phys Chem Chem Phys*, 2008, **10**:4573
- 34) Sabolovic J. *et al.*, *Inorg Chem*, 2003, **42**:2268
- 35) Adamo C. *et al.* *J Chem Phys*, 1999, **111**:2889
- 36) Klamt A., *J Phys Chem A*, 1995, **99**:2224
- 37) Klamt A., *J Phys Chem A*, 1996, **100**:3349
- 38) Klamt A and Schüürmann G., *J Chem Soc Perkin Trans*, 1993, **2**:799
- 39) Hallahan T.W., *Proc Natl Acad Sci USA*, 1991, **85**:5061

5. Bistability and microdomains in cell fate decision

“If you'd like to know, I can tell you that in your universe you move freely in three dimensions that you call space. You move in a straight line in a fourth, which you call time, and stay rooted to one place in a fifth, which is the first fundamental of probability. After that it gets a bit complicated, and there's all sort of stuff going on in dimensions thirteen to twenty-two that you really wouldn't want to know about.”

Mostly harmless – D.Adams

5.1 The biological question

After dealing with investigation on molecular structures we started wondering if and how the inner structure of a molecule could induce a biological effect at a higher than molecular level of complexity.

We could rely for that on a rich literature background since the interpretation of biological regulatory systems in mathematical and statistical terms has gained growing relevance in the last decades. The opportunity of reconstruct and simulate models of biological systems both from a topology and a dynamic point of view has progressively allowed a finer and finer understanding of some complex regulatory behaviors such as bistability, switching, buzzing and blinking [1]. Our purpose was to test if a correlation between a regulatory motif and the molecular structure of own of its component could exist.

Because of that, we decided to focus on the bistable regulating systems of a fundamental biological system as the Ras/Sos signaling pathway activation.

Membrane-bounded Ras protein can cycle between an inactive (GDP-bound) and an active (GTP-bound) states. Its regulatory mechanism is controlled by GTPase activating proteins, (e.g.RasGAP), which hydrolyzes GTP back to GDP and guanine nucleotide exchange factors (e.g.Sos), which binds RasGDP and promotes the exchanging of GDP with GTP.

Several studies showed that the catalytic activity of Sos is strongly determined by its inner structure: the activation of its catalytic site is strongly induced by the present of a second allosteric Ras binding site (Fig.1)[2]. Binding of RasGDP to this allosteric pocket results in a 5-fold increase in GEF activity, whereas binding of RasGTP involves a 75-fold increase [3,4,5]. Thus, SOS-mediated Ras activation involves positive feedback regulation by its own catalytic product, RasGTP (Fig1).

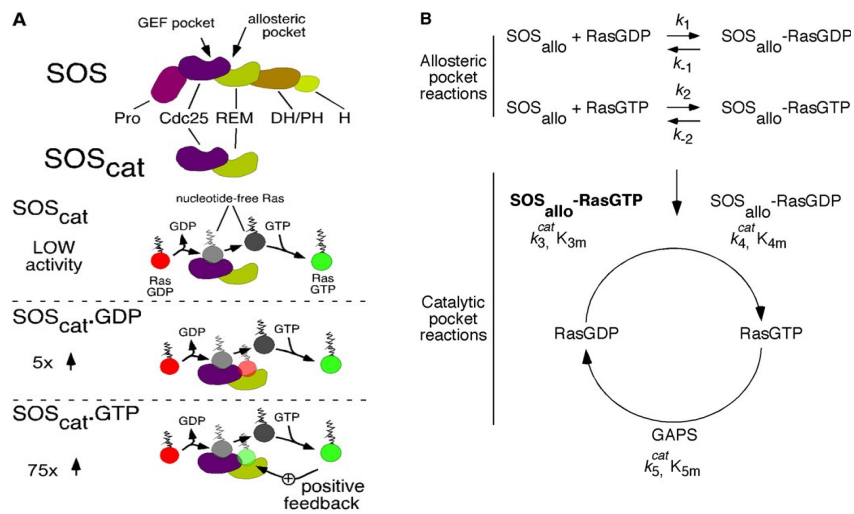


Fig.1. A. The catalytic and allosteric sites of Sos and their reaction with RasGDP and RasGTP. In the lower part the positive feedback is represented. B. The catalytic and allosteric activation cycle of RasGTP. [5]

A computational and experimental work by Das *et al.* [5] reported that the SOS-dependent Ras activation in lymphocytes, beyond stimulus, causes a

digital (*on/off*) response that implies a positive feedback loop in the regulation mechanism of SOS activity. The deterministic model developed by Das et al. demonstrates that for low or high levels of SOS there is only one possible state correlated to low or high levels of active Ras, respectively. For intermediate values of SOS, three states of Ras activity are generated (two states are stable and could be simultaneously observed, the third is unstable and slightly perturbed). Bistability occurs since system lies in the lower state until this is no longer possible and then jumps to an high Ras activity correlated state [5].

An interconnected work reproduced the same biological system in a minimal 2D lattice model in order to investigate the relevance of the microdomains clustering in the signaling pathway. It was reported that positive feedback, in the presence of slow diffusion, results in clustering of activated molecules on the plasma membrane and rapid spatial spreading as the front of the cluster of the newly formed membrane bounded Ras-GTP propagates with a constant velocity dependent on the feedback strength [6].

The role of microdomains in cell signaling has been object of intense study in recent years.

It has been demonstrated that the plasma membrane is made of a lipid bilayer that contains sphingolipids and cholesterol organized in microdomains termed lipid rafts [8]. The lipid rafts microdomains have been shown to play a fundamental role in signaling the cells through the engagement of T cell receptors (TCR), B cell receptors such as I ϵ R, growth factor receptors, chemokine, and interleukins receptors as well as the insulin receptor [9,10]. The outer leaflet of lipid rafts is composed of both sphingolipids with highly saturated acyl chains that pack tightly into gel-like microdomains and cholesterol. The sphingolipids promote formation of a liquid-ordered phase. The sphingolipids and cholesterol are also partitioned

out of the glycerophospholipid bilayer in a liquid-disordered phase due to the unsaturated acyl chains of glycerophospholipids [11]. This phase is termed non-lipid rafts. [8]

The ability of lipid rafts to segregate protein receptors provides a mechanism for compartmentalization of signaling components in the plasma membrane, concentrating certain components in lipid rafts and excluding others.

The proteins within the rafts microdomains are associated with the outer leaflet of rafts through glycosylphosphatidylinositol (GPI)-linkage [12]. It has been postulated that the raft association of GPI-linked proteins is critical to their ability to signal [12], although the mechanism is not completely understood. Cytoplasmic proteins also associate with the inner leaflet of lipid rafts through acylation. Proteins that are dually acylated by saturated fatty acids (*N*-myristoylation and *S*-palmitoylation) partition into the rafts, while those modified by unsaturated fatty acids or prenyl groups are excluded [10]and [13].

The lipid rafts are estimated to represent 40% of the immune cell membranes, as determined by fluorescence anisotropy measurements [14]. In resting T cells, the rafts appear as submicroscopic structures (50 nm in diameter) containing a high number of lipids and a small number of proteins. These structures are referred to as elemental rafts [15,16]. Upon cross-linking of various protein receptors on T cells, the lipid rafts microdomains become larger than 100 nm in diameter, more stable structures, and often attached to the actin cytoskeleton. A scheme of the mechanism of T cells activation through lipid rafts is depicted in fig 2.

In this part of the thesis we decide to investigate the bistability of the Ras/Sos signaling proper of T Cell activation and to test wheter a correlation between this behavior and the clustering of molecules in microdomains

within the cell plasmatic membrane could exist. In order to do this, we utilized brownian dynamic stochastic simulations and highly defined spatial resolution. The final goal was in fact both to develop a model more similar to the biological conditions.

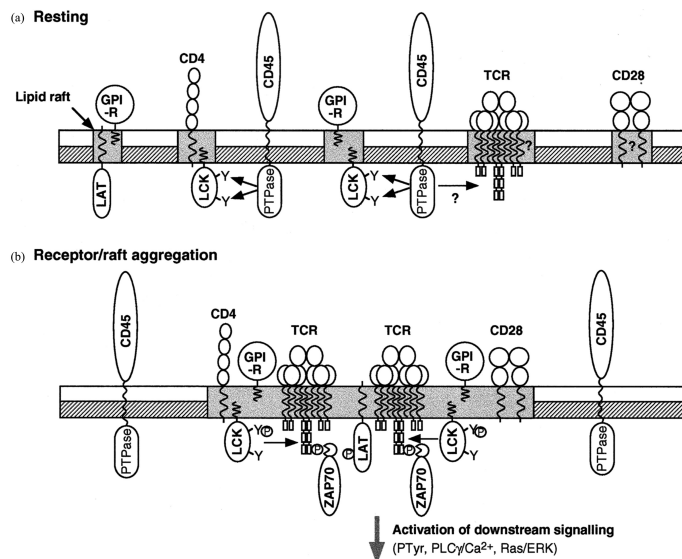


Figure 2. Model for T cell activation via lipid raft aggregation. a. The resting T cell contains small, dynamic lipid rafts, each with a limited number of associated molecules, probably including the TCR and CD28. Due to the small size of the rafts, CD45, although not raft-associated, is able to dephosphorylate both regulatory tyrosines on LCK, with the net effect of inhibiting its kinase activity. CD45 may also further inhibit signalling by acting on the TCR ITAMs or other LCK substrates. b. Ligation of TCRs by an APC results in aggregation of lipid rafts and associated signalling molecules, while excluding CD45, thus promoting LCK activity, ZAP-70 recruitment and subsequent tyrosine phosphorylation. Signalling can similarly be induced by antibody cross-linking of the TCR, and enhanced by co-crosslinking with co-receptors like CD28. Furthermore, these events can be mimicked to some extent by cross-linking GPI-anchored receptors or by direct aggregation of rafts by CT-B cross-linking of GM1. [17]

5.2. Theoretical approach

The most accurate way to depict the time evolution of a system of chemically reacting molecules is to do a molecular dynamics simulation,

meticulously tracking the positions and velocities of all the molecules, and changing the populations of the species appropriately whenever a chemical reaction occurs. Chemical reactions in this approach are viewed as distinct, essentially instantaneous physical events, and they are generally of two elemental types: unimolecular, occurring as a result of processes internal to a single molecule, and bimolecular, occurring as a result of a collision between two molecules.

From a classical mechanics viewpoint, one might suppose that such a system is deterministic, in that a given initial state always leads to the same state at some specified later time. But this is not the case: first of all, even if the system evolved deterministically with respect to the positions, velocities, and molecular populations of the species, it would not evolve deterministically with respect to the species populations alone. Secondly, quantum indeterminacy unavoidably enters because in a unimolecular reaction we can never know exactly when a molecule will transform itself into a different isomeric form. Third, chemical systems are usually not mechanically isolated since they are in contact with a heat bath, whose essentially random perturbations keep the system in thermal equilibrium at some temperature.

In view of the fact that molecular populations in a chemically reacting system are integer variables that evolve stochastically, it is remarkable that chemical kinetics has traditionally been analyzed using a mathematical formalism in which continuous (real) variables evolve deterministically: Traditional chemical kinetics holds that in a well-stirred, thermally equilibrated chemical system, the number of molecules X_i of each chemical species S_i ($i = 1, \dots, N$) evolves in time according to a set of coupled ordinary differential equations (ODEs) of the form:

$$dX_i/dt = f_i(X_1, \dots, X_N) \quad (i = 1, \dots, N)$$

where the functions f_i are inferred from the specifics of the various reactions. This set of equations is usually expressed in terms of the concentration variables $Z_i \equiv X_i/\Omega$, where Ω is the system volume. Unluckily when the system is small enough that the molecular populations of at least some of the reactant species are not too many orders of magnitude larger than one, discreteness and stochasticity may play important roles. Whenever that happens, and it often does in cellular systems in biology, ODEs turns out to be not the best choice to mathematically describe the system's behavior. Stochastic chemical kinetics attempts to describe the time evolution of a well stirred chemically reacting system in a way that takes honest account of the system's discreteness and stochasticity.

5.2.1 Stochasticity

If we consider a well-stirred system of molecules of N chemical species $\{S_1, \dots, S_N\}$, which interact through M chemical reactions $\{R_1, \dots, R_M\}$. It is assumed that the system is confined to a constant volume and is in thermal (but not chemical) equilibrium at some constant temperature.

$X_i(t)$ denote the number of molecules of species S_i in the system at time t . The goal is to estimate the state vector

$$\mathbf{X}(t) \equiv (X_1(t), \dots, X_N(t)),$$

given that the system was in state $\mathbf{X}(t_0) = \mathbf{x}_0$ at some initial time t_0 .

Because of the conditions responsible for the system to be well stirred, the system's state can be described by specifying only the molecular populations, ignoring the positions and velocities of the individual

molecules. The fundamental assumption being made is that the overwhelming majority of molecular collisions that take place in the system are elastic (nonreactive), and further that the net effect of these elastic collisions is twofold: first, the positions of the molecules become uniformly randomized throughout; second, the velocities of the molecules become thermally randomized to the Maxwell-Boltzmann distribution.

To the extent that this happens, is it possible to ignore the nonreactive molecular collisions, whose simulation would have a high computation and time cost.

The changes in the species populations are of course a consequence of the chemical reactions. Each reaction channel R_j is characterized mathematically by two quantities.

The first is its state-change vector

$$\mathbf{v}_j \equiv (v_{1j}, \dots, v_{Nj}),$$

where v_{ij} is the change in the S_i molecular population caused by one R_j reaction, so if the system is in state \mathbf{x} and one R_j reaction occurs, the system immediately jumps to state $\mathbf{x} + \mathbf{v}_j$. The other characterizing quantity for R_j is its propensity function a_j , which is defined so that $a_j(\mathbf{x})dt$ corresponds to the probability, given $\mathbf{X}(t) = \mathbf{x}$, that one R_j reaction will occur somewhere inside in the next infinitesimal time interval $[t, t + dt)$.

If R_j is the unimolecular reaction $S_1 \rightarrow \text{product(s)}$, the underlying physics dictates the existence of some constant c_j , such that $c_j dt$ gives the probability that any particular S_1 molecule will so react in the next infinitesimal time dt . It then follows from the laws of probability that if there are currently x_1 S_1 molecules in the system, the probability that some one of them will undergo the R_j reaction in the next dt is $x_1 \cdot c_j dt$. Thus the propensity function is

$$a_j(\mathbf{x}) = c_j x_1.$$

If R_j is a bimolecular reaction of the form $S_1 + S_2 \rightarrow \text{product(s)}$, kinetic theory arguments and the well-stirred condition together imply the existence of a constant c_j , such that $c_j dt$ gives the probability that a randomly chosen pair of S_1 and S_2 molecules will react according to R_j in the next infinitesimal time dt . The probability that some one of the $x_1 x_2$ S_1 - S_2 pairs inside Ω will react according to R_j in the next dt is therefore $x_1 x_2 \cdot c_j dt$, so that the propensity function is $a_j(\mathbf{x}) = c_j x_1 x_2$. If instead the bimolecular reaction had been $S_1 + S_1 \rightarrow \text{product(s)}$, the number of distinct S_1 molecular pairs as $\frac{1}{2} x_1(x_1 - 1)$, and so obtained for the propensity function $a_j(\mathbf{x}) = c_j \frac{1}{2} x_1(x_1 - 1)$.

Evaluating c_j completely from first principles is a challenging task, requiring specific assumptions to be made about how the reaction R_j physically occurs. Although the probabilistic nature of the propensity function precludes making an exact prediction of $\mathbf{X}(t)$, the probability could be inferred as follow:

$$P(\mathbf{x}, t | \mathbf{x}_0, t_0) = \text{Prob} \{ \mathbf{X}(t) = \mathbf{x}, \text{ given } \mathbf{X}(t_0) = \mathbf{x}_0 \}.$$

A time-evolution equation for $P(\mathbf{x}, t | \mathbf{x}_0, t_0)$ by applying the laws of probability to the propensity law could be derived, bringing to the *chemical master equation* (CME) (10–12):

$$\frac{\partial P(\mathbf{x}, t | \mathbf{x}_0, t_0)}{\partial t} = \sum_{j=1}^M [a_j(\mathbf{x} - \mathbf{v}_j) P(\mathbf{x} - \mathbf{v}_j, t | \mathbf{x}_0, t_0) - a_j(\mathbf{x}) P(\mathbf{x}, t | \mathbf{x}_0, t_0)]$$

In principle, the CME completely determines the function $P(\mathbf{x}, t | \mathbf{x}_0, t_0)$.

But the CME is actually a set of coupled ODEs, with one equation for every possible combination of reactant molecules. It is therefore not surprising that the CME can be solved analytically for only a few simple cases, and even

numerical solutions are prohibitively difficult in other cases.

So the key to generating simulated trajectories of $\mathbf{X}(t)$ is not the function $P(\mathbf{x}, t | \mathbf{x}_0, t_0)$, but rather a new probability function $p(\tau, j | \mathbf{x}, t)$, which is defined as follows:

$p(\tau, j | \mathbf{x}, t) d\tau$ = the probability, given $\mathbf{X}(t) = \mathbf{x}$, that the next reaction in the system will occur in the infinitesimal time interval $[t + \tau, t + \tau + d\tau)$, and will be an R_j reaction.

Formally, this function is the joint probability density function of the two random variables time to the next reaction (τ) and index of the next reaction (j), given that the system is currently in state \mathbf{x} . It is not difficult to derive an exact formula for $p(\tau, j | \mathbf{x}, t)$ by applying the laws of probability to the fundamental premise. The result is:

$$p(\tau, j | \mathbf{x}, t) = a_j(\mathbf{x}) \exp(-a_0(\mathbf{x})\tau),$$

where

$$a_0(\mathbf{x}) \triangleq \sum_{j'=1}^M a_{j'}(\mathbf{x})$$

This equation is the mathematical basis for the stochastic simulation approach. It implies that τ is an exponential random variable with mean (and standard deviation) $1/a_0(\mathbf{x})$, while j is a statistically independent integer random variable with point probabilities $a_j(\mathbf{x})/a_0(\mathbf{x})$. There are several exact Monte Carlo procedures for generating samples of τ and j according to these distributions.

Since a systematic description of all of them and of all the stochastic algorithms generated in the last decades is far beyond the aim of this thesis, a brief description of the one used for this project is presented in the next

section.

5.2.2 *Smoldyn* software

Smoldyn software has been chosen because it allows to simulate a chemical reaction network with a spatial resolution that is accurate to nearly the size scale of individual molecules. It has been developed by Steven S Andrews and Dennis Bray of the University of Cambridge, UK [7].

In this software each molecule is treated as a point-like particle that diffuses freely in three-dimensional space. When a pair of reactive molecules collide, such as an enzyme and its substrate, a reaction occurs and the simulated reactants are replaced by products. For bimolecular reactions, the inputs to the simulation are experimental reaction rates, diffusion coefficients and the simulation time step. From these are calculated the simulation parameters, including the ‘binding radius’ (separation for a molecular collision) and the ‘unbinding radius’ (initial separation between a pair of reaction products). In the following section few methodological details characterizing the software are presented.

5.2.2.1 *Basic theory*

The theoretical basis behind simulations in *Smoldyn* is an extension of the *Smoluchowski model* for *diffusion-influenced* systems [18].

In the model, time increases continuously, as it does in nature, but in contrast to the finite time steps that are introduced in the next section for the simulation algorithms. Each molecule is treated as a point-like particle that

diffuses freely in space with continuously variable x , y and z coordinates, quantified with Fick's laws [19]:

$$\mathbf{J}_B(\mathbf{r}, t) = -D_B \nabla \rho_B(\mathbf{r}, t)$$

$$\rho_B(\mathbf{r}, t) = D_B \nabla^2 \rho_B(\mathbf{r}, t)$$

B is some generic chemical species, $\mathbf{J}_B(\mathbf{r}, t)$ is the flux of B molecules at position \mathbf{r} and time t , $\rho_B(\mathbf{r}, t)$ is the local number concentration of B molecules, and D_B is the diffusion coefficient for B [20].

The coordinates of a molecule are its center of mass. The Smoluchowski description also accounts for external and long-range forces (such as between ionic species [21]) but they're ignored in the program since they have minimal influence in a typical biochemical system and they are computationally expensive to simulate. To allow the use of Fick's laws on small size scales as well as large ones, the dynamics of the solvent and other unreactive species are ignored [22], leading to infinitely detailed *Brownian motion* of the reactive molecules. This approximation makes the results only accurate on size scales that are somewhat larger than those of individual molecules.

Steric interactions are ignored between molecules that do not react with each other, which is valid for dilute solutions.

Molecular spatial orientations and internal energy levels typically fluctuate on time scales that are faster than the diffusive and reactive processes that are of interest [21, 23], allowing them to be ignored as well. Because of these approximations, the complete time-dependent state of the model is fully specified by a list of the molecular positions.

By definition, a *diffusion-limited* bimolecular reaction occurs very rapidly once two reactive molecules come into contact, which happens when the molecular centers are separated by a distance equal to the sum of the molecular radii.

This description is used for the Smoluchowski model in which a bimolecular reaction occurs at the moment when two reactive molecules collide with each other. Smoluchowski derived the steady-state reaction rate for this physical description, in terms of the molecular radii and the diffusion coefficients. However, most reactions occur at a slower rate because of a reaction activation energy. This is addressed in the Smoluchowski model by replacing the sum of the molecular radii with a smaller effective *binding radius* (σb), thus yielding the correct steady-state reaction rate for all bimolecular reactions, regardless of the reaction mechanism.

Reversible reactions, such as the generic reaction $A + B \leftrightarrow C$, pose a problem. If the A and B products of the backward reaction are initially separated by the binding radius, which is the obvious separation, then the ensuing Brownian motion of A and B makes them almost certain to collide again. This leads to a nearly instantaneous reaction back to C (the actual probability for recollision is 1 and the expected time that elapses before reacting is 0) [17]. This recollision problem is not addressed in the Smoluchowski model because it does not consider reversible reactions. In order to solve this Andrews et al. developed a Smoluchowski-related model: the A and B dissociation products are initially separated by a fixed distance which is larger than σb , called the *unbinding radius* (σu) [19]. Using this rule, neither inter-molecular forces nor reaction probabilities need to be introduced, leaving diffusion as the sole fundamental process. After unbinding, the A and B product molecules may diffuse away from each other or they may diffuse together again and rebind, called a *geminat recombination* [21, 27].

An unbinding radius is an artificial concept but its use can be justified. Physically, a C molecule is an A–B complex, for which the interaction potential energy is a function of the A–B separation, typically with an

activation barrier [28].

If diffusion influences the system even a small amount, any boundary between reactants and products may be crossed many times. To prevent this, it is helpful to introduce bistability by defining a boundary on each side of the activation barrier: a forward reaction occurs when the A–B separation is less than the inner boundary and a reverse reaction occurs when it is greater than the outer boundary [29]. The Smoldyn model does not have an activation barrier, although the two boundaries are considered.

The model becomes quantitatively accurate when it is supplemented with experimental data, including diffusion coefficients and reaction rates.

5.2.2.2 *The algorithm*

To develop a simulation of a system the continuous time of the model is replaced with steps of length t , which can be kept constant throughout the simulation [30] (which is done in *Smoldyn*) or made adaptive. The end of each step is like an observation of a virtual system that evolves continuously. In particular, molecules are considered to move with infinitely detailed Brownian motion, even though the detail is neither explicitly simulated nor observable. Using this interpretation, these algorithms are designed to yield observable results that are as close as possible to the analytically derived dynamics of the model system.

The simulation errors can be made arbitrarily small because the simulated dynamics become identical to those of the model in the limit of small time steps.

The program framework used in Smoldyn is summarized as in fig.3.

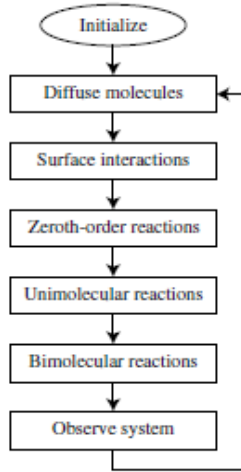


Fig.3 Flowchart of the algorithm implemented in *Smoldyn* [7].

5.2.2.3 Molecular diffusion

Because the model considers individual molecules rather than concentrations, Fick's second law is rewritten as a master equation by replacing the number density of B molecules with the spatial probability density for a single molecule:

$$p_B(\mathbf{r}, t) = D_B \nabla^2 p_B(\mathbf{r}, t)$$

The product $p_B(\mathbf{r}, t)d\mathbf{r}$ is the probability that a specific B molecule is within volume $d\mathbf{r}$ about position \mathbf{r} at time t .

In a simulation, a molecule starts at a known position and diffuses over the course of a time step. Solving the above equation for this initial condition shows that the probability density for the displacement of a molecule after a time step has a Gaussian profile on each Cartesian coordinate [19, 32]:

$$p_B(\mathbf{r} + \Delta\mathbf{r}, t + \Delta t) = G_{s_B}(\Delta x)G_{s_B}(\Delta y)G_{s_B}(\Delta z)$$

$$G_s(\Delta x) \equiv \frac{1}{s\sqrt{2\pi}} \exp\left(-\frac{\Delta x^2}{2s^2}\right)$$

$$s_B \equiv \sqrt{2D_B\Delta t}$$

where x , y and z are the Cartesian displacements, $G_s(x)$ is a normalized Gaussian with mean 0 and standard deviation equal to s , and s_B is the *root mean square (rms) step length* of species B.

These results form the basis of a simulation method called *Brownian dynamics* [32,34] in which diffusion is simulated by picking a normally distributed random displacement for each molecule at each time step.

5.2.2.4. Surfaces

Nearly all physical reaction systems are confined to a finite volume, making it necessary to simulate surface interactions.

Surfaces are most easily treated as arrays of flat panels. From a computational viewpoint, surface types include the following: inert impermeable surfaces, which prevent molecules from passing from one side to the other; periodic boundaries of the simulation volume, which do not exist physically but are useful for the simulation of systems with effectively infinite extent; and absorbing surfaces, which irreversibly capture all molecules that diffuse into them. In each case, the algorithm has to determine whether each molecule interacted with each panel of the surface during the previous time step using the standard criterion that all observable dynamics should be indistinguishable from those of the model.

Impermeable surfaces are considered first. Solving equation $p_B(\mathbf{r}, t) = D_B\nabla^2 p_B(\mathbf{r}, t)$ with an impermeable plane as a boundary condition shows that

the spatial probability density, $p_B(\mathbf{r}, t)$, reflects off the surface like light from a mirror [32]. Thus, diffusion in the presence of inert impermeable surfaces is accurately simulated using ballistic-type reflections [33]. In the algorithm, each molecule is propagated forward over t according to equation

$$p_B(\mathbf{r} + \Delta\mathbf{r}, t + \Delta t) = G_{s_B}(\Delta x)G_{s_B}(\Delta y)G_{s_B}(\Delta z)$$

then, the straight line path of the molecule is reflected off any surface that it crosses.

Periodic boundaries are similar. Because the equation above is correct in the absence of surfaces, it is also correct for periodic boundaries, provided that any probability density that escapes the system is translated across the simulation volume.

In the algorithm, any molecule that diffuses past a boundary is transferred across the system as though it had followed a straight line over the course of the time step.

An absorbing surface is treated by temporarily considering it to be permeable and asking the question: what is the probability that a specific molecule crossed the surface during the time step? If the molecule started on the inside and diffused to the outside, then it obviously crossed the surface and should be absorbed. It could also start and end on the inside but have crossed the surface during the time step, the probability of which can be found using the initial and final perpendicular distances to the surface, denoted by l_i and l_f , respectively (these are positive if the molecule is inside and negative if it is outside). The probability that the molecule crossed the surface at least once, starting from distance l_i , conditioned with the additional knowledge of l_f , is

$$\begin{aligned} \text{Prob}(\text{cross}|l_f) &= 1 - \text{Prob}(\text{no cross}|l_f) \\ &= 1 - \frac{\text{prob}(\text{no cross}, l_f)}{\text{prob}(l_f)} \end{aligned}$$

$$\text{prob}(\text{no cross}, l_f) = G_{s_B}(l_f - l_i) - G_{s_B}(l_f + l_i)$$

$$\text{prob}(l_f) = G_{s_B}(l_f - l_i).$$

Prob() is a probability, prob() is a probability density, a vertical line indicates a conditional probability, and a comma indicates a joint probability [35]. For example, Prob(cross| l_f) is the probability that the boundary is crossed, given a knowledge of l_f , and prob(no cross, l_f) dl is the probability that the boundary is not crossed and the final distance is between l_f and $l_f + dl$. The densities are found with previous equation and the initial condition that the molecule starts at l_i away from a surface [32]. For the joint density, a boundary condition is that the probability that the molecule is at the surface is 0. Results are

$$\text{prob}(\text{no cross}, l_f) = G_{s_B}(l_f - l_i) - G_{s_B}(l_f + l_i)$$

$$\text{prob}(l_f) = G_{s_B}(l_f - l_i).$$

Substituting these into equation yields the desired answer:

$$\text{Prob}(\text{cross}|l_f) = \exp\left(-\frac{2l_i l_f}{s_B^2}\right)$$

Thus, the algorithm for absorbing surfaces is that a molecule should be absorbed if it ends up on the far side of the surface at the end of a time step or if a random number with a uniform distribution between 0 and 1 is less than $\exp(-2l_i l_f / s_B^2)$.

All of these methods are exact for planar surfaces but are in error for curved surfaces or near junctions of flat surface panels. In general, the spatial resolution is approximately the molecular rms step length.

5.2.2.5. Zeroth-order reactions

A zeroth-order reaction progresses at a rate that is independent of all chemical concentrations, implying that product molecules are formed spontaneously. While unphysical, zeroth-order reactions can be useful components of simulations because they can provide chemical inputs to the simulated system without requiring a complete treatment of the input mechanism.

If the product of the reaction is A and k_0 is the rate constant, the zeroth-order mass-action rate law is $\rho_A = k_0$.

On average, $k_0 t$ product molecules are formed during each time step. However, this has some stochastic variation, which is given with a Poisson distribution [35].

5.2.2.6. Unimolecular reactions

Unimolecular reactions are described by the generic equation $A \rightarrow \text{products}$. This might describe a true unimolecular reaction, such as a molecular dissociation, or a bimolecular reaction between an A molecule and an abundant species that is not explicitly simulated.

The kinetics are typically of first order:

$$\rho_A = -k_1 \rho_A$$

where k_1 is the first-order rate constant.

Upon integration, the probability that a specific A molecule reacts during t is

$$\text{Prob}(\text{reaction}) = 1 - \exp(-k_1 t)$$

If an A molecule can react via multiple first-order pathways, a sequential application of the previous equation leads to a bias towards the first pathway that is attempted. Instead, solution of a collection of equations, where the i th reaction has a first-order rate constant $k_{1,i}$, leads to the reaction probabilities [17]:

$$\text{Prob}(\text{reaction } i) = \frac{k_{1,i}}{\sum_j k_{1,j}} \left[1 - \exp\left(-\Delta t \sum_j k_{1,j}\right) \right]$$

5.2.2.7 Bimolecular reactions

Bimolecular reactions, described by the generic equation $A + B \rightarrow C$, have the steady-state reaction rate:

$$\rho_C = k_2 \rho_A \rho_B$$

where k_2 is the second-order rate constant, from which it is possible to find the binding radius of the model.

An exact algorithm would be based on a question similar to that posed above for absorbing surfaces: given the positions of molecules A and B before and after a time step, what is the probability that the distance between them was less than the binding radius at some point during the time step? Unlike the other algorithms, an exact solution for bimolecular reactions is not practical. So the authors chose an algorithm that is intuitive, simple and very fast: two molecules always react if they end up within $\sigma_b(t)$ at the end of a time step and never react if the final separation is greater than that. For reversible reactions, dissociation products are initially separated by $\sigma_u(t)$. These parameters are analogous to the binding and unbinding radii of the model system and approach them in the limit of small time steps.

The correct binding and unbinding radius are derived from a combinatorial

process that involve simulations and experimental results

5.3 Results

5.3.1. Ras-Sos activation signaling three-dimensional model

The first step to investigate the biological problem was to implement the model proposed in [6] using *Smoldyn*.

The 2D-lattice model proposed by the authors was challenging but not completely satisfying from a physiological point of view. This step was also use for testing the program for our next three-dimensional model.

We firstly developed a two-dimensional model that reproduced the Das model using the same parameters of the reference work [6] in order to test the efficiency of the chosen methodology. The 2D model we obtained showed evidences of clustering and spreading of RasGTP during the simulations (data not shown).

Benchmarked *Smoldyn* software, we decided that we could add another layer of complexity to the model and to developed a three-dimensional model using the parameters presented in the work of Das et al [5].

A $4 \mu\text{m}^2 \times 20 \text{ nm}$ box was created, using $[\text{Ras-GDP}]_i = 47 \text{ molec}/\mu\text{m}^2$ (membrane), $[\text{Sos}]_i = 60 \text{ molec}/\mu\text{m}^3$ (in *solution*), $[\text{Ras-GAP}]_i = 20 \text{ molec}/\mu\text{m}^3$ and diffusion coefficients of $D_{\text{Ras}} = 0.1 \mu\text{m}^2/\text{s}$ and $D_{\text{SOS}} = D_{\text{RasGAP}} = 1 \mu\text{m}^2/\text{s}$ [5]. RasGDP is confined in a compartment of $4 \mu\text{m}^2 \times 1.7 \text{ nm}$ one of the wider side of the box whose height corresponds to the estimated radius of gyration of RasGDP (Fig.4).

The list of the considered reactions is the following:

- $\text{SOS} + \text{Ras-GDP} \rightarrow \text{SOS-Ras-GDP}$ (k1)
- $\text{SOS-Ras-GDP} \rightarrow \text{Ras-GDP} + \text{SOS}$ (k-1)
- $\text{SOS} + \text{Ras-GTP} \rightarrow \text{SOS-Ras-GTP}$ (k2)
- $\text{SOS-Ras-GTP} \rightarrow \text{Ras-GTP} + \text{SOS}$ (k-2)
- $\text{SOS-Ras-GDP} + \text{Ras-GDP} \rightarrow \text{SOS-Ras-GDP-Ras-GDP}$ (k3)
- $\text{SOS-Ras-GDP-Ras-GDP} \rightarrow \text{SOS-Ras-GDP} + \text{Ras-GDP}$ (k-3)
- $\text{SOS-Ras-GDP-Ras-GDP} \rightarrow \text{SOS-Ras-GDP} + \text{Ras-GTP}$ (k3c)
- $\text{SOS-Ras-GTP} + \text{Ras-GDP} \rightarrow \text{SOS-Ras-GTP-Ras-GDP}$ (k4)
- $\text{SOS-Ras-GTP-Ras-GDP} \rightarrow \text{SOS-Ras-GTP} + \text{Ras-GDP}$ (k-4)
- $\text{SOS-Ras-GTP-Ras-GDP} \rightarrow \text{SOS-Ras-GTP} + \text{Ras-GTP}$ (k4c)
- $\text{GAP} + \text{Ras-GTP} \rightarrow \text{GAP-Ras-GTP}$ (k5)
- $\text{GAP-Ras-GTP} \rightarrow \text{GAP} + \text{Ras-GTP}$ (k-5)
- $\text{GAP-Ras-GTP} \rightarrow \text{GAP} + \text{Ras-GDP}$ (k5c)

The associated reaction rates were: $k_1 = 0.12 \mu\text{M/s}$, $k_{-1} = 3.0 \text{ 1/s}$, $k_2 = 0.11 \mu\text{M/s}$, $k_{-2} = 0.4 \text{ 1/s}$, $k_3 = 0.07 \mu\text{M/s}$, $k_{-3} = 1.0 \text{ 1/s}$, $k_{3c} = 0.003 \text{ 1/s}$, $k_4 = 0.05 \mu\text{M/s}$, $k_{-4} = 0.1 \text{ 1/s}$, $k_{4c} = 0.038 \text{ 1/s}$, $k_5 = 1.74 \mu\text{M/s}$, $k_{-5} = 0.2 \text{ 1/s}$, $k_{5c} = 0.1 \text{ 1/s}$.

Equations 1,-1, 2,-2 describe the binding between the allosteric site of Sos to RasGDP or RasGTP. Equation 3, -3, 3c describe the catalytic activity of Sos-RasGDP complex, while equations 4, -4, 4c describe the catalytic activity of Sos-RasGTP complex, which determines the feedback loop, while equations 5, -5, 5c describe the RasGAP activity.

The simulations were carried out for 180 s, using time steps of 1ns.

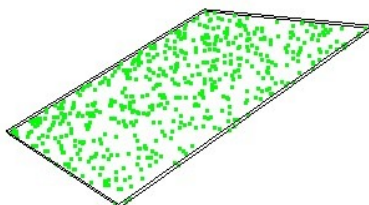


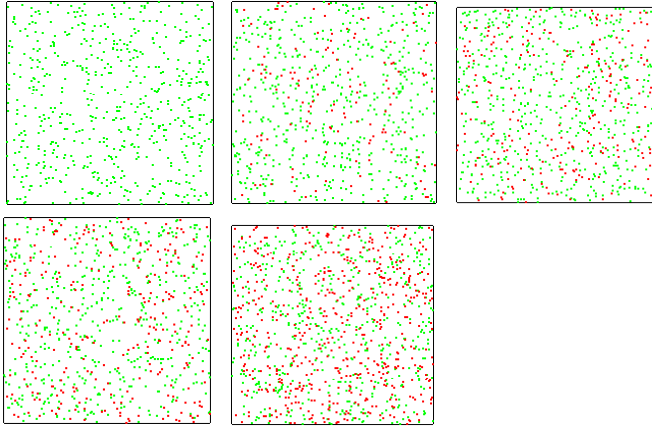
Fig.4. The 3D box utilized for the simulations. Sos is depicted in green.

In fig.5 results from simulations are shown. It is evident the formation of clusters along the surface when the positive feedback loop is active (*middle*). Otherwise, a uniform distribution of RasGTP along the surface is displayed (*top*).

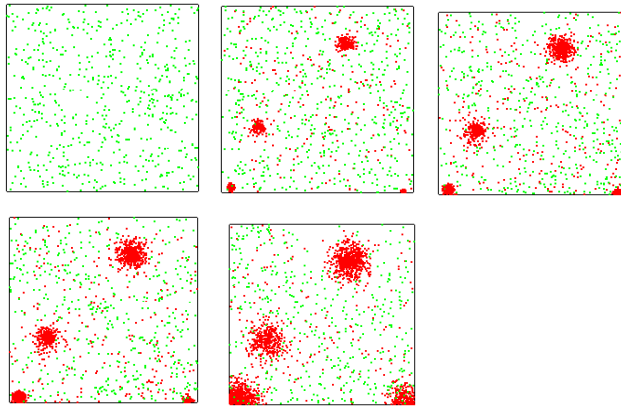
The positive feedback loop determined by the allosteric site of Sos induces bistability in the system, as it is shown in fig.5.(bottom). As was reported in the reference paper for low and high level of Sos only one state is present, while for concentration of Sos within this interval, three states are observed, two stable and one unstable states.

Our model is a significant step forward more physiological likely modeling [6]: all the reactions of the biological systems have been taken into account and the brownian motion of the molecules allows to observe realistic kinetic and catalytic mechanisms.

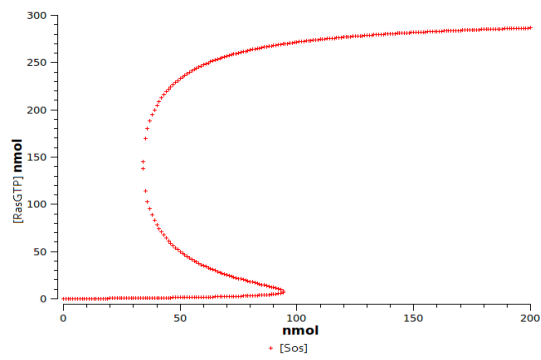
These kind of results could be considered a proof that the methodology we used constitutes a plausible way to study this kind of spatial-related biological problems.



Top



Middle



Bottom

Fig. 5. Snapshots of the spatial simulation, taken at different times. *Top:* Simulation of the system without the positive feedback; a) 0s, b) 45s, c) 90s, d) 135s, f) 180s. *Middle:* Simulation of the system with the positive feedback activated; a) 0s, b) 45s, c) 90s, d) 135s, f) 180s. RasGTP displayed in red, SOS displayed in green. Bottom: Graphic representation of the steady states reached by the system along the simulation.

5.3.2. Analogy between fermions and bistable systems

In the last few years a new perspective to the study of spatial contribution to cell behavior has been suggested. Lipshtat *et al.* [36] proposed an analogy between fermions and biological bistable systems.

Fermions are particles that obey the *Pauli Exclusion Principles* that a quantum state can be either empty or occupied by a single particle. In the same way, GTPases are associated with molecules (GTP or GDP), where each copy of GTPase can be found in one of two possible states. The distribution of fermions in different states is determined by the Fermi–Dirac distribution according to which the probability that a particle is in the state of energy E_i is a function of energy level and temperature:

$$P(E_i) = (1/Z) \exp(-E_i/k_B T)$$

Z is the partition constant, K_B is Boltzmann's constant (R/N_A), T is the temperature and E_i is the energy of the state i .

Following the same logic the activation level of a GTPase could be described as the probability of finding a GTPase molecule in the GTP-bound state. In this case the probability is exponentially dependent on the signal ($[X]$). So signal could play for GTPases as the same role as energy plays for fermions. It results that GTPase activation follows Fermi–Dirac distribution

as function of $[X]$ [36].

These new considerations could be fundamental to get new insight in the study of spatial gradients (and consequentially of lipid rafts).

Spatial gradients in fact show exponential dependence from the signal and from the source so that they bring to the formation of intracellular regions of high activity molecules (microdomains). If there is a point source at one side of the cell, and one component spreads out by diffusion, the concentration of the component decreases exponentially with the distance from the source formation of multiple biochemical compartments without physical boundaries.

In order to find an optimal quantitative description of the distribution into different states and of the spreading in our model, this new theoretical approach has lately being taken into consideration and it will mainly constitute the new phase of study in this project.

5.4. Conclusions

The model we developed provides a new insight on the bistability and microdomains clustering in biological systems.

Our work represents also an example of the results that could be obtained applying brownian dynamic stochastic simulations to the study of biological spatial issues.

Further improvements of this project will particularly regard quantitative and statistical studies of the bistable behavior. The hypothesis of the application of the Fermi-Dirac distribution to biological systems will also be tested.

5.5. References

- 1) Tyson *et al.*, *Curr.Op.Cell.Biol*, 2003, **15**:221-231
- 2) Margarit,S.M. *et al.*, *Cell*,2003, **112**:685–695
- 3) Freedman T.S. *et al.*, *Proc. Natl. Acad. Sci.*, 2006 **103**, 16692–16697
- 4) Sondermann H. *et al.*, *Cell*, 2004, **119**:393–405
- 5) Das J. *et al.*, *Cell*, 2009 **136**:337-351
- 6) Das J. *et al.* *Chem.Phys*, 2009, **130**:245102-10
- 7) Andrews S.S. and Bray D., *Phys.Biol*, 2000, **1**:137-151
- 8) Sunil T. *et al.*, *Molecular Immunology*, 2004, **41**:399–409
- 9) Simons and Toomre, *Nat. Rev. Mol. Cell Biol.*, 2000, **1**:31–41
- 10) Bromley *et al.*, *Annu. Rev. Immunol.*, 2001, **19**:375–396
- 11) Brown and London, *Annu. Rev. Cell Dev. Biol.*, 1998, **14**:111–136
- 12) Horejsi *et al.*, *Immunol. Today*, 1999, **20**: 356–361
- 13) Melkonian *et al.*, *J. Biol. Chem.*, 1999, **274**:3910–3917
- 14) Gidwani A. *et al.* *Biochemistry*, 2001, **40**:12422–12429
- 15) Varma R. *et al.*, *Nature*, 1998, **394**:798–801
- 16) Pralle A., *J. Cell Biol.*, 2000, **148**:997–1007
- 17) James P., *Seminars in IMMUNOLOGY*, 2000, 12:3-34
- 18) von Smoluchowski M. V. Z. *Phys. Chem.*,1917, **92**: 129
- 19) Berg H. C. *Random Walks in Biology* 2nd edn 1993 (Princeton, NJ: Princeton University Press)
- 20) Elowitz M. B. *et al.* *J. Bacteriol.* 1999, **181**:197
- 21) Rice S. A. *Diffusion Limited Reactions (Comprehensive Chemical Kinetics* vol 25 ed C H Bamford, C F H Tipper and R G Compton (Amsterdam: Elsevier), 1985
- 22) Schnell S. and Turner T. E. *Prog. Biophys. Mol. Biol.* 2004, **85**:235
- 23) Northrup S. H. and Erickson H. P. *Proc. Natl Acad. Sci. USA*, 1999, **89**:3338
- 24) Collins F. C. and Kimball G. E. *J. Colloid Sci.* 1949, **4**:425
- 25) Agmon N. *J. Chem. Phys.* 1984, **81**:2811
- 26) Berlin Y. A. *et al.*, *J. Chem. Phys.* 1980, **73**:4619
- 27) Noyes R. M. *J. Am. Chem. Soc.* 1995, **77**:2042
- 28) Atkins P. W. *Physical Chemistry 1986* 3rd edn (New York:Freeman)
- 29) Northrup S H and Hynes J T 1980 *J. Chem. Phys.* **73** 2700
- 30) Crank J, *The Mathematics of Diffusion* 2nd edn (Oxford: Oxford University Press), 1975
- 31) Edelman A. L. and Agmon N., *J. Chem. Phys.* 1993 **99**:5396
- 32) Ermak D. L. and McCammon J. A., *J. Chem. Phys.* 1978, **69** 1352
- 33) Snell J. L., *Introduction to Probability* (New York:Random House), 1988
- 34) Lipshtat A., *PNAS*, 2010, **107**:1247–1252

6. Dynamic model of the G1/S phase transition in the $G_{q/11}$ protein signaling network

“If you try and take a cat apart to see how it works, the first thing you have on your hands is a non-working cat”
Quoted by Richard Dawkins - Eulogy for Douglas Adams

6.1. The biological question

In the third and last part of the thesis we step onto a higher level of complexity and we investigated cellular-level phenomena.

We decided to focus our attention on the mammalian cell cycle and in particular on the G1/S phase transition. The $G_{q/11}$ signaling pathway was chosen as the starting point of our investigation since it is involved in different fundamental physiological processes as sensory perception, behavioral and mood regulation, regulation of the immune systems activity. Dysfunction in this pathway can lead to different type of cancer.[1]

The biological questions we have been trying to answer were: What are the dynamics in the cellular pathways that lead to the G1/S transition? How a dysfunction may affect this pathways? In order to do this, it would be desirable to build a model suitable for predictions whose construction would consist of the following step:

- Build a dynamical model with all the known components and interactions of the system
- Put in evidence which are the main regulator motifs in the model that control the proliferation events
- Perturb the system according to gene expression data of cancerous

cells to identify which are the critical points in the regulatory network

- Use the model for predictions in order to identify possible therapeutic treatment for certain type of cancer (in particular, small cells lung cancer and brain cancer)

In this part of the thesis we developed the first step and we built a complete dynamic model of the system.

Many hormones, neurotransmitters and sensory stimuli elicit cellular response through the activation of seven transmembrane receptors coupled to heterotrimeric G proteins that consist of three subunits α , β and γ . The GTPase cycle of heterotrimeric G proteins is driven between inactive and active forms by binding to GTP and its hydrolysis [2, 3]. Activation of the G protein-coupled receptors (GPCRs) by extracellular stimuli induces the release of GDP and binding of GTP on $G\alpha$ protein subunit. This leads to the dissociation of $G\alpha\beta\gamma$ into $G\alpha$ -GTP and $G\beta\gamma$. $G\alpha$ -GTP and $G\beta\gamma$ then transmit the receptor-generated signals to downstream effector molecules until the intrinsic GTPase activity of G hydrolyzes GTP to GDP and the inactive $G\alpha$ -GDP reassociates with $G\beta\gamma$. G proteins are divided according to structural and functional similarities of their G into four groups; G_s , G_i , G_q , and G_{12} [3]. $G\alpha_q$ members including $G\alpha_q$, $G\alpha_{11}$, $G\alpha_{14}$, and $G\alpha_{15/16}$ (mouse/human orthologues, respectively) stimulate the β -isoforms of phospholipase C (PLC- β) that catalyzes the hydrolysis of phosphatidylinositol bisphosphate, resulting in the generation of inositol trisphosphate (IP_3) and diacylglycerol (DAG). These second messengers serve to propagate and amplify the Gq-mediated signal with intracellular calcium mobilization by Ca^{2+} release from IP_3 -regulated intracellular stores, and DAG-dependent protein kinase C (PKC) activation. Ca^{2+} and PKC participate in diverse signaling to evoke

different cellular events [4].

In our model we decided to study in particular all the events that draw proliferation. To accomplish this we need to add various macro-components to the ones just indicated. The MAP kinases cascade has thus been taken into account, in particular the phosphorylation level of ERK, JNK, p38. At nuclear level, we included the activation of both early and delayed expressed genes, as c-jun, c-fos, c-myc as well as the genes that are directly involved in the G1/S transition, as cycD, cycE and cycA. (Fig1). To describe this components the best, we need to consider seven different compartments: extracellular space, plasma membrane, cytosol, nucleus, nuclear membrane, endoplasmatic reticulum and its membrane.

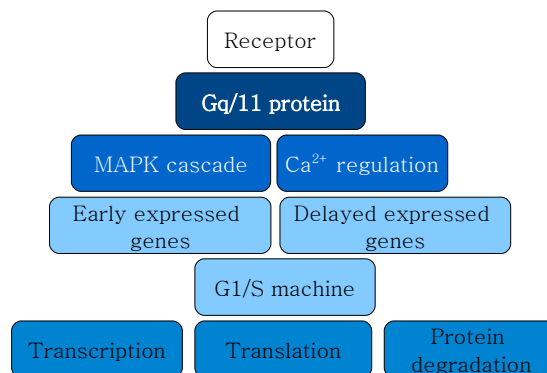


Fig1. The macro components of our network.

Here a brief description of the components on how we decide to deal with the main components:

- Receptor: We chose for both our experiment and calculations to adopt a new generation class of receptors called *Receptor activated solely by a synthetic ligand* (RASSL). This receptor is an engineered version of a human muscarinic m3 receptor that allows precise

spatial and temporal control of the signaling *in vivo*. (Fig.2). Experimental data furnished by S.Gutkind and collaborators (Oral and Pharyngeal Cancer Branch, National Institute of Dental and Craniofacial Research, National Institutes of Health, Bethesda, Maryland, USA).

- MAPK cascades and Ca^{2+} regulation: simulation parameters were adapted from a previous published work [5]
- Gene expression: data were partially furnished by other published models in literature, in particular: Novak (2004) [6], Haberichter (2007) [7], Conradie (2010)[8], Yao (2008), Aguda (1999) [9] and from experimental data furnished by Gutkind *et al.*
- Transcription, translation, protein degradation: data were adapted from previous works [5,7].

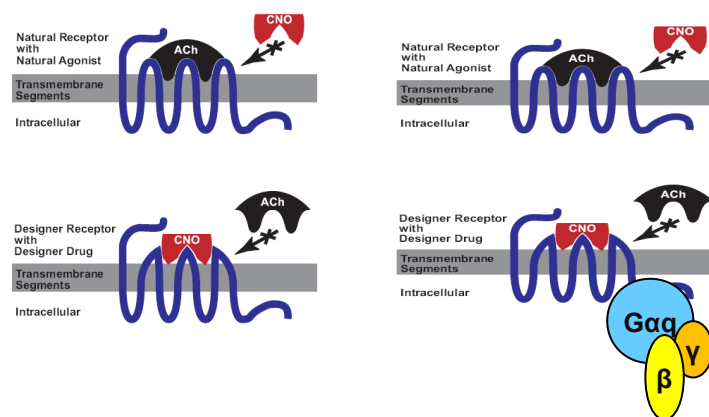


Fig.2 The construction of a designer receptor. Native hM_3 receptor responds to the natural hormone (ACh) but does not respond to CNO (*Left*). This receptor is matured through saturation mutagenesis to result in a designer receptor (*Right*), which responds only to the designer drug (CNO) and is impervious to the natural hormone. Adapted from [10].

All the components chosen, the system was firstly displayed in a topological network in order to put into evidence all the proteins and reactions that need

6.2. Theoretical approach

6.2.1. Ordinary differential equations

For this kind of study we choose the application of *deterministic algorithms*. This algorithm, given a particular input, will always produce the same output, and the underlying machine will always pass through the same sequence of states. It computes a mathematical function; a function has a unique value for any given input, and the algorithm is a process that produces this particular value as output.

Temporal dynamics are described adequately by ordinary differential equations (ODE), the most common type of representation used to model cell signaling and metabolic pathways in cases when the cell can be considered as a well-stirred reactor and when stochastic effects can be neglected.

An ODE is an equality involving a function and its derivatives. An ODE of order n is an equation of the form

$$F(x, y, y', \dots, y^{(n)}) = 0$$

where y is a function of x , $y' = dy/dx$ is the first derivative with respect to x , and $y^{(n)} = d^n y/dx^n$ is the n th derivative with respect to x .

Nonhomogeneous ordinary differential equations can be solved if the general solution to the homogenous version is known, in which case the undetermined coefficients method or variation of parameters can be used to find the particular solution.

An ODE of order n is said to be linear if it is of the form

$$a_n(x)y^{(n)} + a_{n-1}(x)y^{(n-1)} + \dots + a_1(x)y' + a_0(x)y = Q(x)$$

A linear ODE where $Q(x)=0$ is said to be homogeneous.

In general, an n th-order ODE has n linearly independent solutions. Furthermore, any linear combination of linearly independent functions solutions is also a solution.

Simple theories exist for first-order (integrating factor) and second-order (Sturm-Liouville theory) ordinary differential equations, and arbitrary ODEs with linear constant coefficients can be solved when they are of certain factorable forms. Integral transforms such as the Laplace transform can also be used to solve classes of linear ODEs. Morse and Feshbach give canonical forms and solutions for second-order ordinary differential equations.

While there are many general techniques for analytically solving classes of ODEs, the only practical solution technique for complicated equations is to use numerical methods. The most popular of these is the Runge-Kutta method, but many others have been developed, including the collocation method and Galerkin method. A vast amount of research and huge numbers of publications have been devoted to the numerical solution of differential equations, both ordinary and partial (PDEs) as a result of their importance in fields as diverse as physics, engineering, economics, and electronics.

The solutions to an ODE satisfy existence and uniqueness properties. These can be formally established by Picard's existence theorem for certain classes of ODEs. Let a system of first-order ODE be given by

$$dx_i/dt = f_i(x_1, \dots, x_n, t)$$

for $i=1, \dots, n$ and let the functions $f_i(x_1, \dots, x_n, t)$, where $i=1, \dots, n$, all be defined in a domain D of the $(n+1)$ -dimensional space of the variables x_1, \dots, x_n, t . Let these functions be continuous in D and have continuous first partial derivatives $\partial f_i / \partial x_j$ for $i=1, \dots, n$ and $j=1, \dots, n$ in D . Let (x_1^0, \dots, x_n^0) be in D .

Then there exists a solution of the first-order ODE given by

$$x_1 = x_1(t), \dots, x_n = x_n(t)$$

for $(t_0 - \delta) < t < (t_0 + \delta)$ (where $\delta > 0$) satisfying the initial conditions

$$x_1(t_0) = x_1^0, \dots, x_n(t_0) = x_n^0$$

Furthermore, the solution is unique, so that if

$$x_1 = x_1^*(t), \dots, x_n = x_n^*(t)$$

is a second solution for $(t_0 - \delta) < t < (t_0 + \delta)$, then $x_i(t) \equiv x_i^*(t)$ for $(t_0 - \delta) < t < (t_0 + \delta)$. Because every n th-order ODE can be expressed as a system of n first-order ODEs, this theorem also applies to the single n th-order ODE.

6.2.2. *Virtual Cell*

The *Virtual Cell Modeling and Simulation Framework* has been developed by The National Resource for Cell Analysis and Modeling. This software platform has been designed to model cell biological processes. This new technology associates biochemical and electrophysiological data describing individual reactions with experimental microscopic image data describing their subcellular locations.

Cell physiological events can then be simulated within the empirically derived geometries, thus facilitating the direct comparison of model predictions with experiment.

The *Virtual Cell* consists of a biological and mathematical framework. Scientists can create biological models from which the software will generate the mathematical code needed to run simulations.

The simulations are run over the internet on 84 servers with 256 GHz total

CPU power and 119 GB total RAM.

6.2.2.1. CVODE

Algorithm CVODE has been chosen to solve the set of ODEs of the model. CVODE is a solver for stiff and nonstiff initial value problems for systems of ordinary differential equation (ODEs)

An ODE initial value problem can be written as

$$y' = f(t, y), y(t_0) = y_0, \quad y \in \mathbf{R}^N$$

where y' denotes the derivative dy/dt . This problem is stiff if it contains one or more strongly damped modes (i.e. the Jacobian matrix $J = \partial f / \partial y$ has an eigenvalue with large negative real part).

The underlying integration methods used in CVODE are variable-coefficient forms of the Adams and BDF (Backward Differentiation Formula) methods. The numerical solution to the ODE equation is generated as discrete values y_n at time points t_n . The computed values y_n obey a linear multistep formula

$$\sum_{i=0}^{K_1} \alpha_{n,i} y_{n-i} + h_n \sum_{i=0}^{K_2} \beta_{n,i} \dot{y}_{n-i} = 0$$

Here the y'_n are computed approximations to $y'(t_n)$, $h_n = t_n - t_{n-1}$ is the stepsize, and $\alpha_{n,0} = -1$. For use on nonstiff problems, the Adams-Moulton formula is characterized by $K_1 = 1$ and $K_2 = q$, and the order q varies between 1 and 12. For stiff problems, the BDF formula has $K_1 = q$ and $K_2 = 0$, and the order q varies between 1 and 5. In either case, the nonlinear system

$$G(y_n) \equiv y_n - h_n \beta_{n,0} f(t_n, y_n) - a_n = 0,$$

where

$$a_n \equiv \sum_{i>0} (\alpha_{n,i} y_{n-i} + h_n \beta_{n,i} \dot{y}_{n-i})$$

must be solved (approximately) at each time step. In the nonstiff case, this is usually done with simple functional (or fixed point) iteration. In the stiff case, the solution of is usually done with some variant of Newton iteration. This requires the solution of linear systems of the form

$$M[y_{n(m+1)} - y_{n(m)}] = -G(y_{n(m)})$$

where M is an approximation to the Newton matrix $I - h\beta_{n,\rho}J$ and $J = \partial f/\partial y$ is the ODE system Jacobian. The linear systems arising in the Newton iteration are solved by either a direct or an iterative method. The direct solvers currently available in CVODE include dense (full) and banded methods, and a diagonal approximate Jacobian method. The iterative method is the Generalized Minimal Residual method, or GMRES. With scaling and preconditioning included, this method is defined SPGMR.

The estimation and control of errors is an extremely important feature of any ODE solver. In CVODE, local truncation errors in the computed values of y are estimated, and the solver will control the vector e of estimated local errors in accordance with a combination of input relative and absolute tolerances. Specifically, the vector e is made to satisfy an inequality of the form

$$\|e\|_{WRMS,ewt} \leq I$$

where the weighted root-mean-square norm $\|e\|_{WRMS,w}$ with weight vector w is defined as

$$\|v\|_{WRMS,w} = \sqrt{\frac{\sum_{i=1}^N (v_i w_i)^2}{N}}$$

The CVODE error weight vector ewt has components

$$ewt_i = [1/(RTOL \cdot |y_i| + ATOL_i)]$$

where the non-negative relative and absolute tolerances $RTOL$ and $ATOL$ are specified by the user. $RTOL$ is a scalar, but $ATOL$ can be either a scalar or a

vector. The local error test controls the estimated local error e_i in component y_i in such a way that, roughly, $|e_i|$ will be less than $ewt_i^{-1} = RTOL \cdot |y_i| + ATOL_i$. The local error test passes if in each component the absolute error $|e_i|$ is less than or equal to $ATOL_i$, or the relative error $|e_i|/|y_i|$ is less than or equal to $RTOL$. Using of $RTOL = 0$ is required for pure absolute error control, and $ATOL_i = 0$ for pure relative error control.

6.2.2.2. *Parameter estimation*

Parameter estimation in the Virtual Cell provides tools to optimize parameters in compartmental, i.e. non-spatial, model to best fit experimental data.

The optimization algorithm CFSQP (C code for Feasible Sequential Quadratic Programming) as been used for the calculations. SQP is an iterative method for nonlinear optimization and is used on problems for which the objective function and the constraints are twice continuously differentiable. SQP methods solve a sequence of optimization subproblems, each which optimizes a quadratic model of the objective subject to a linearization of the constraints. If the problem is unconstrained, then the method reduces to Newton's method for finding a point where the gradient of the objective vanishes. If the problem has only equality constraints, then the method is equivalent to applying Newton's method to the first-order optimality conditions, (the so-called Karush–Kuhn–Tucker conditions, OR KKT) of the problem.

6.2.2.3. *Experimental data*

Data have being provided by Silvio Gutkind *et al.* (Oral and Pharyngeal

Cancer Branch, National Institute of Dental and Craniofacial Research, National Institutes of Health, Bethesda, Maryland, USA).

Experiments were conducted using a NIH3T3-derived cell line stably expressing RASSL-G_q.

So far, MAPKinases activation and gene expression profiles were detected.

Experimental data are reserved since they are going to be imminently published by S.Gutkind and collaborators.

6.2.2.4. Methodological issues

Implementing our model in VirtualCell has not been free from applicational issues. The main difficulty was that not all the sub-pathways of the system are univocally determined.

To overcome this uncertainties and to choose reliable components for our model, the “Occam razor principle” has been applied. According to it “*entities should not be multiplied unnecessarily*”, i.e. when you have two competing theories that make exactly the same predictions, the simpler one is the better. In our case when divergent opinions on a phenomenon were provided in literature we chose the simplest combination of components and interactions that could explain it.

The second issue we had to face was the combinatorial explosion of the number of molecules and reactions once we implemented the model in *Virtual Cell*. First of all for almost the totality of the molecules involved both the activated and inactivated form has to be taken into account; secondarily the presence of different compartments implies another duplication in the number of those molecules that undergo exportation from or recruitment to a

compartment. Because of this, the total number of both reactions and molecules easily arose above 200 and the total number of parameters (reaction rates) reached more than 600 units .

This lead to the most challenging part of the implementation. Since only a low percentage of this parameters can be found in literature and since also these data have been achieved from different *in vitro* experiments so that they're rarely comparable, we need to substantially rely on parameter estimation. Fortunately, the large number of estimated parameters does not necessarily mean that the model is flexible enough to generate any type of result. The specific form of the equations themselves restricts the possible result. Still, in the future, different parameter combinations will have to be tested systematically in order for the system-level models to become truly predictive. Very recently, a number of groups started to make promising progress in this direction by using genetic and stochastic optimization algorithms to automatically explore the whole model parameter space [12]. Such efforts require a significant amount of quantitative data to be used to “score” the model’s success, so modeling efforts become inseparable from experimental ones.

6.2.2.5. Simulations settings

Compartment	Volume-Area
Cytosol	1000 μm^3
Plasma membrane	483,52 μm^2
Nucleus	100 μm^3
Nuclear membrane	104,17 μm^2
Endoplasmatic reticulum	50 μm^3
ER membrane	65,62 μm^2
Extracellular space	3000 μm^3

Simulation settings	
Algorithm	CVODE
Max timestep	1s
Output every	300 s
Rel tolerance	1.0E-9
Abs tolerance	1.0E-9
Start time	0
End time	50400

The complete list of the reaction parameters and the initial concentrations values can be found in Virtual Cell, username: Valentina_B, model

6.3. Results

We developed a multi compartment ordinary differential equation model in *Virtual Cell* and numerical simulations have been carried out to obtain the time courses of RASSL dependent activation of GTPases, MAP-kinases and different genes.

6.3.1. Activation of MAPKs

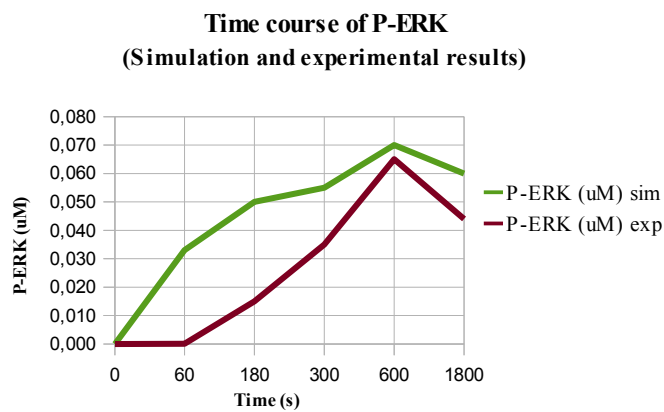
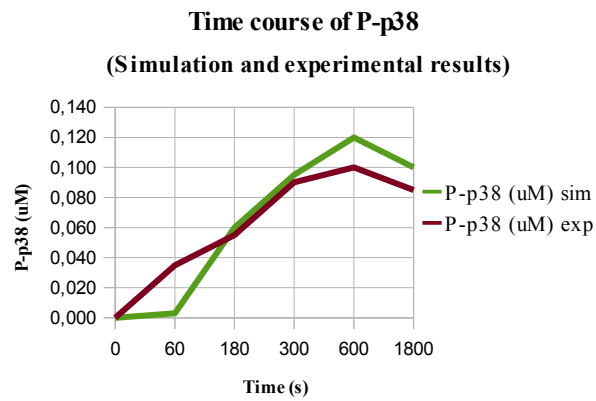
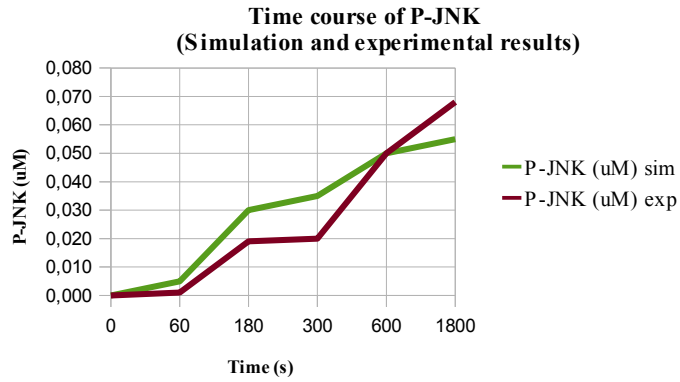


Fig.4. Time courses of MAPKinases JNK, ERK, p38. In *green* the simulation results and in *red* the experimental results.

The time courses for the activation of JNK, p38 and ERK obtained from the simulations (Fig.4) reasonably fit with the experimental data provided by S. Gutkind et al..

6.3.2. Activation of c-jun and c-fos gene

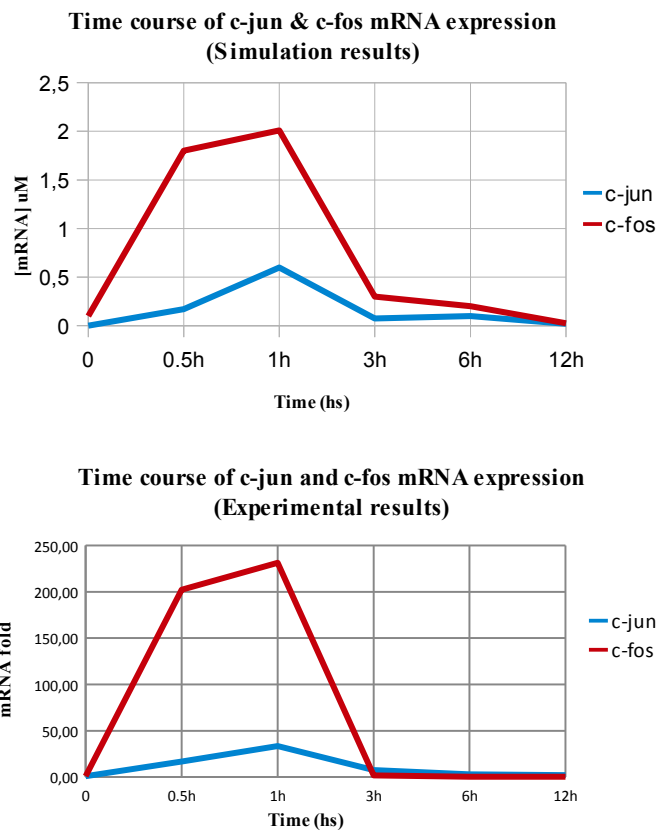
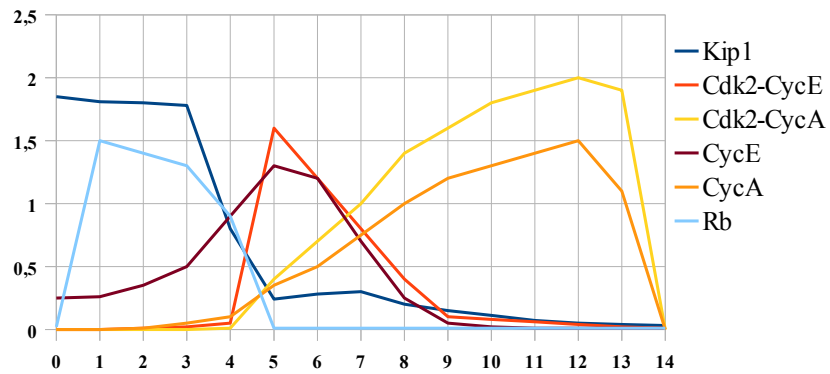


Fig.5. Time courses of c-jun and c-fos MRNA expression. Above, the simulation results; bottom. The experimental results. C-jun in blue, c-fos in red .

The time course for the expression of c-jun mRNA expression obtained from the simulation reasonably fits with the experimental data provided by S. Gutkind et al. The highest level of expression is reached after 1h.

6.3.3. Activation of some components of the G1/S regulating machine

Time course of G1/S transition components
(Novák, Tyson, 2004)



Time course of G1/S transition components
Simulation results

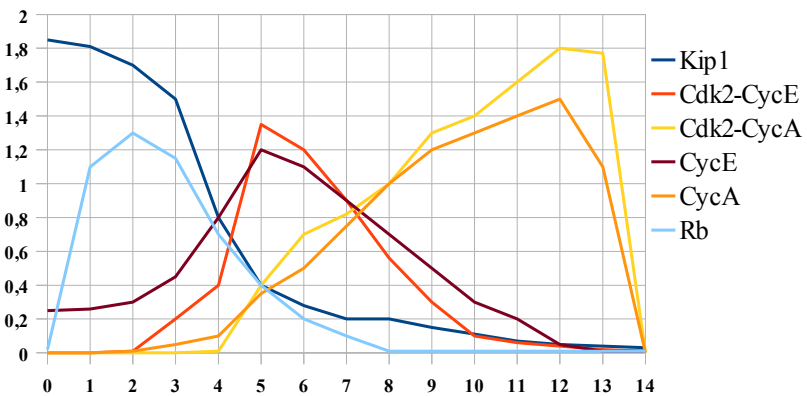


Fig.5. Time courses of some of the components of the G1/S transition machine. *Above*, elaboration of the results obtained by Novak and Tyson [6]. *Bottom*, the simulation results. Kip1 (*blue*), Cdk2-CycE (*red*) Cdk2-CycA (*yellow*), CycE (*violet*), CycA (*orange*), Rb (*light blue*).

The time course for the expression of the main components of the G1/S transition machine obtained from the simulation fits with the model of Novak and Tyson [6].

Accordingly to this results after 3-4hs the Rb level are low enough to allow cycE and cycA transcription and translation so that they can form complexes with Cdk2.

The levels of expression of other components of this part of the system need to be taken into account and they would be part of the object of our future investigation

6.4 Conclusions and perspectives

Considering the complexity and the size of the pathway in study, the results obtained so far are encouraging for future developments of the model.

The overall goal of the present work will be to identify the pivotal points of the regulation of the network that could represent pharmaceutical targets in cancer treatments, the current model will need further constraints refinements through new experimental data and the addition of other modules, as the cytoskeleton regulation.

Simulations will be carried out to study the influence of perturbations on the activation and expression of the components of the network in order to simulate a cancerous event.

In order to overcome the huge amount of parameters in the model, a

correlation analysis between reaction parameters will also be carried out.

A dynamical graph will be finally designed matching the numerical data obtained from the *Virtual Cell* simulations and the interaction and influence graphs obtained by a Petri net model, in order to provide a clearer picture of the dynamic of the system.

6.5. References

- 1) Dorsam and Gutkind, *Nature Reviews Cancer*, 2007, **7**: 79-94
- 2) Kaziro Y *et al. Annu Rev Biochem*, 1991, **60**: 349–400
- 3) Simon *et al.*, *Science* 1991, **252**: 802–808.
- 4) Mizuno and Itoh, *Neurosignals*, 2009; **17**:42–54
- 5) Bhalla and Iyengar, 1999, *Science* **283**, 381-387
- 6) Novák and Tyson, *J theor biol*, 2004, **230**:563:579
- 7) Haberichter *et al*, *Molec Syst Biol*, 2007, **3**:1-8
- 8) Conradie *et al*, *FEBS*, 2010, **277**, 357:367
- 9) Yao *et al*, *Nat cell biol*, 2008, **10**:476:482
- 10) Aguda and Tang, *Cell prolifer*, 1999, **32**:321-335
- 11) Armbruster *et al.*, *Proc Natl Acad Sci USA* (2007), **104**:5163–5168
- 12) Mezer *et al*. *Biochim biophys Acta*, 2006,**1763**:345-355.

7. Final conclusions

*I rarely end up where I was intending to go, but often I end
up somewhere I needed to be.*

D.Adams

The aim of the thesis is the development of computational models of the enzymatic activity of biomolecules at different scales.

To briefly resume the achievements of this thesis, it can be asserted that all the studies lead to very satisfying results, even though the stochastic and systemic project certainly need further investigations and improvements.

With the DFT studies we have been able to kinetically and thermodynamically characterize the reaction mechanism of a unusual molecule, amavadin, and its thiolic substrate MMA. This could help to discover its biological function. Our study of the binding between angiogenin and copper would also hopefully help to give some insight into the life-threatening angiogenic process and to maybe develop therapeutics agents in the future.

Studying bistable behaviors using stochastic processes allowed us to try to combine two different theoretical approaches to answer one scientific question. The promising results we obtained will push us to further investigate these methodologies as well as new applications of them.

Finally, dealing with a cellular-size model gave us a true glimpse on the complex systems and all the fascinating behavior that emerge from them. In our opinion this kind of models will be more and more useful in the future to understand human biology and physiology. In order to do this, methodologies need to be refined, in particular to deal with issues related to the huge number of parameters.

In conclusion, the models we proposed may constitute a reliable demonstration of the importance of quantitative modeling in biology. Our work also suggests that even if all the approaches we used were almost mathematically orthogonal to each other, they share a common methodology. This is not only evident in such cases as in of stochasticity, where the theoretical basis draws on Quantum Chemistry principles, but also because in each of these approaches we had to deal with the same issues of convergence, minimization, optimization, etc. This implies that an underlying correlation running through all the scales of modeling can be hypothesized.

8. Acknowledgments

“So long, and thanks for all the fish”

The hitchhiker's guide to the galaxy - D.Adams

First of all I'd like to thank prof. Piercarlo Fantucci to welcome me in his laboratory and because he supported me when I decide to swithc to another route.

A special thought also for Prof. Luca De Gioia who really indicated me that route and who shared with me all the challenges of this thesis.

I'm also very grateful to Dott. Luca Bertini and Dott.Giuseppe Zampella who drove me through my first paper, classical music and the world of Miyazaki.

A heartfelt thanks to Prof. Ravi Iyengar, who welcomed me into his laboratory that really played a decisive turn in my life. Thanks for all you taught me and for your support through hard times. Another person who deserves a bow is Simon Hardy, my supervisor during my staying in Ravi's lab: thanks for all your teachings and for being so patient with me. All the best for you and your family.

A sincere thanks also to Prof. Silvio Gutkind and Jose Vaque for their precious collaboration in the cell cycle project.

I don't have words to describe how much I own to my beloved parents who gave me all they have to make me achieve this goal. This work is not really

sufficient to repay you for all the love and support you gave me during these years. I'll try my best to do it in the future.

Thanks also to all my family for their support, most of all in this last year.

I wouldn't have got through these three years without my Paul: we really had quite a hectic time but you never let me down even when I didn't deserve it. I'm so happy that we're climbing the same route again.

With a big hug I'd like to thanks all my beloved friends: Bashi e Manu who make me feel always at home, even when I've been away for months; Nella, my treasure flatmate and one of the most good-hearted person I've ever known; the other components of the "Fantastic 4": Cris, Moreno, Manu. I wish you all the best for the future (and the conquer of the Universe); Eleonora and Elena for being Minnie and HelloE; Lorenzo e Jacopo because they've been there since the beginning.

Another bear hug goes to all the guys of the Molecular Modeling lab: Ale, because sometimes allows me to be Napalm, Stefano, my adventure mate in the jungle of stochasticity, Claudio, for our talks on the most disparate subjects, Elena, for her support and enthusiasm, Matteo the restless man, Alberto and his "Vale...", ILLambru, Matteo because of D.Adams, Andrea because he surfs, Giulia R. for being caring, Giulia F.(you should have come to our lab before!). Also thanks to the other guys who used to work in our lab: Francesca, Rubben, Ale Dido, the two Marco P, Dimitrios. All the best also to Francisco, matchless flatmate!

A special thanks to Chiara and Massi (and Talpo) for adopting me during my year in NY, I would have never survived the American life style without your

help. Thanks also for helping me during hard times. I will probably never be able to explain you how much it meant to me.

Bernadette, you would have deserved much more than my frequent lunatic mood, but I was and am really happy to know you. Thanks for supporting me through all that happened, for our cheesy talks and for your inspiring attitude towards life.

A special thought also is for Sherry, my personal Emily, for standing me during the major crisis with my “spaghetti model” and for sharing with me (and with tons of laugh) up and downs of life.

Thanks to all the other guys of the Iyengar Lab: Arjun, Jens, Gomathi, Amrita, Shan, Evren, Julia, Nikos, Frank, Jo and all the others for this amazing year I spent in you lab and in NY.

I'm also very grateful to Renny, Miriam, Luz, Kyoko for helping me getting through bureaucratic and administration stuff and for their support.

Finally a heartfelt “Grazie!” and “un abbraccio” to Regina, whose friendship was one of the most precious thing I found in NY.

9. Publications

“If there's anything more important than my ego around, I want it caught and shot now”

The hitchhiker's guide to the galaxy – D.Adams.

Copper coordination to the putative cell binding site of angiogenin. A DFT investigation.

Luca Bertini,^a Maurizio Bruschi,^b Marco Romaniello,^a Giuseppe Zampella,^a Matteo Tiberti,^a Valentina Barbieri,^a Claudio Greco,^a Diego La Mendola,^c Raffaele P. Bonomo,^d Piercarlo Fantucci,^a Luca De Gioia.^a

a) Dipartimento di Biotecnologie e Bioscienze, Università degli Studi di Milano Bicocca, Piazza della Scienza, 2 – 20126 Milano (Italy).

b) Dipartimento di Scienze Ambientali, Università degli Studi di Milano Bicocca, Piazza della Scienza, 1 – 20126 Milano (Italy).

c) Istituto di Biostrutture e Bioimmagini – C. N. R. – Catania, Viale A. Doria 6, 95125, Catania (Italy).

d) Dipartimento di Scienze Chimiche, Università degli Studi di Catania, Viale A. Doria 6, 95125, Catania (Italy).

Keywords: DFT, bioinorganic chemistry, copper, angiogenin.

Inorg.Chemistry. *Accepted*

Dedicated to Professor Enzo Barone on the occasion of his 60th birthday.

Abstract

We present a DFT study of the structural and spectroscopic properties of the complex formed by Cu²⁺ with the peptide fragment Ac-PHREN-NH₂, which encompasses the putative cell binding domain of angiogenin, as well as with its Ac-PHRQN-NH₂ variant. Analysis of structures, energies and spectroscopic parameters has allowed to conclude that the metal coordination environment at pH 8 is formed by a nitrogen atom of His, two deprotonated amide groups, and an oxygen atom from the COO⁻ side chain of Glu, in nice agreement with recent experimental results [La Mendola D, Magri A, Vagliasindi LI, Hansson O, Bonomo RP, Rizzarelli E (2010) Dalton Trans, 39:10678]. Moreover, DFT results allowed to reveal that the Glu sidechain of the

Ac-PHREN-NH₂ peptide is coordinated in equatorial position, in a tetrahedrally distorted square planar arrangement, fully disclosing the effects of Cu²⁺ binding on the structural properties of this key angiogenin portion. In the Ac-PHRQN-NH₂ variant the carboxylate group is replaced by a H₂O molecule in a coordination arrangement similar to that of the wild type system.

1. Introduction

Angiogenesis is a set of functional processes responsible for the formation of new blood vessels from existing ones, and is also a key component of the homeostatic process that regulates the distribution of oxygen to tissues [1]. Angiogenesis occurs in several tightly regulated stages that orchestrate a network of cooperative interactions, and can be divided in several phases: (1) an initiation phase, characterized by increased vasopermeability; (2) a progression phase, in which proteolytic enzymes that degrade the extracellular matrix and promote endothelial cell migration are produced; (3) a final phase, in which differentiation into new vessels takes place. The latter phase is mediated by molecules that recruit mesenchymal cells to vessel walls [2].

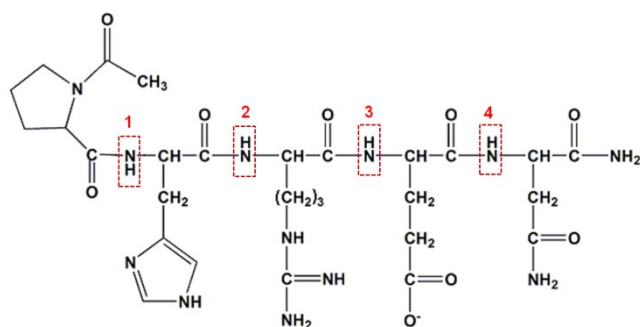
Angiogenin (Ang) is a protein that undergoes nuclear translocation in endothelial cells, where it accumulates in the nucleolus and stimulates rRNA transcription, a rate limiting step in ribosome biogenesis, protein translation, and cell growth [3]. Ang was originally isolated from the conditioned medium of HT-29 human colon adenocarcinoma cells [4]. The mature Ang is a basic, single-chain protein containing 123 amino acids with a molecular weight of about 14.400 Da, and it is a homologue of bovine pancreatic ribonuclease A. Indeed, its ribonucleolytic activity is rather low, but Ang is essential for angiogenesis and other functions [5]. In fact, Ang belongs to the ribonuclease superfamily, showing 33% sequence identity to the pancreatic ribonuclease A. However, although the crystal structures of human Ang and pancreatic ribonuclease A are highly similar, a notable difference is evident in the ribonucleolytic active site: the pyrimidine binding site of Ang is “obstructed” by the Gln117 residue, explaining its very weak ribonucleolytic activity (about 10⁵–10⁶ lower than that of RNase A).

Besides its ribonucleolytic activity, binding of Ang to the endothelial cells surface is needed for its biological functions, and amino acid residues from 60 to 68 are critical in this process. Notably, affinity of Ang for endothelial cells is largely increased in the presence of copper ions [6]. In fact, copper has been recognized to be an angiogenic factor, but the mechanism whereby it exerts this function is still not well understood [7, 8]. In particular, Ang binds 2.4 mol of copper per mol of protein, and the metal ion is important for Ang binding to calf pulmonary artery endothelial cells, which increases by 4.3 fold in the presence of Cu^{2+} .

A putative endothelial cell binding domain of Ang is located in a loop exposed to the solvent and largely unstructured in the native protein [9]. The related amino acid sequence (hAng60–68 = KNGNPHREN) contains the prolyl-histidyl (PH) dyad which is reminiscent of that present in other copper binding proteins, such as prion proteins [10] and the Wilson's and Menkes' ATPases [11]. In this respect, the comparison between Ang and RNase A structures already suggested possible copper binding sites [12]. In fact, RNase A binds copper at several sites including His-12, His-105 and His-119 at pH 5 [13], but Ang features three additional histidine residues that are not present in RNase A: His-8, His-65 and His-84 [14, 15]. The similarity between the copper binding plasma tripeptide Gly-His-Lys and the copper binding site of albumin and α -fetoprotein, where copper binds to a histidyl residue adjacent to a basic residue (Arg or Lys), suggests that His-65, which is adjacent to Arg-66 in Ang, might be involved in copper binding.

Very recently La Mendola et al. [16] reported the synthesis and characterization of the complex formed by the peptide fragments encompassing the sequence hAng64–68 (Ac-PHREN-NH₂; Scheme 1) and the whole sequence hAng60–68 (Ac-KNGNPHREN-NH₂) with copper.

Scheme 1. Structure of the Ac-PHREN-NH₂ peptide. The four NH group of the peptide bonds have been highlighted in red boxes.



In particular, combined potentiometric and spectroscopic investigations allowed to reveal the species distribution and the coordination environments of the Cu(II) complexes. It turned out that both peptides coordinate Cu²⁺ in a similar fashion. Moreover, thermodynamic and spectroscopic data indicated that the side chains of Glu and His residues are involved in copper binding at physiological pH. The Cu(II) interaction with the peptide fragment Ang64–68(E67Q) (Ac-PHRQN-NH₂), in which glutamate was substituted by a glutamine residue, was also studied in order to unveil the role of glutamate carboxylate group on Cu(II) coordination. The comparison between results obtained studying the Cu(II) complexes formed by Ac-PHREN-NH₂ and its E67Q variant provided further evidence of the presence of a carboxylate oxygen atom in the copper coordination sphere. On the ground of such results it was concluded that at pH 8 the metal coordination environment in the complex of Cu(II) with Ac-PHREN-NH₂ is formed by a nitrogen atom of His, two deprotonated amide groups, a water molecule and an oxygen atom from the COO⁻ side chain of Glu. However, it was not possible to reveal whether the carboxylate oxygen atom was coordinated in equatorial or apical position, hindering the full disclosure of the effects of Cu(II) binding on the structural properties of the Ang60-68 protein portion.

With the aim of complementing available experimental data in the elucidation of the nature of the copper first coordination environment, as well as the apical/equatorial disposition of oxygen ligands, we have carried out a DFT investigation of the Cu(II) complexes formed by Ac-PHREN-NH₂, as well as by the single point mutated peptide Ac-PHRQN-NH₂.

2. Computational details

The generation of the starting structures for DFT optimization was carried out using a combined MM/MD approach. In particular, starting structures, differing for coordination geometry and/or orientation of aminoacid side chains, were initially optimized using the MMFF94x forcefield [17, 18] as implemented in the MOE suite (MOE Molecular Operating Environment, version 2008.10; Chemical Computing Group Inc.: Montreal, Canada, 2008). Then, short 1000 ps molecular dynamic simulations were performed at 100K to relax the systems and allow it to escape from metastable high-energy minima. Simulation temperature was kept very low because the aim of such MM/MD simulations was local (and not global) sampling of the potential energy hypersurface.

Quantum chemical geometry optimizations were carried out using the pure Generalized Gradient Approximation (GGA) BP86 DFT functional [19, 20] and the Resolution of Identity (RI) technique [21], as implemented in the TURBOMOLE suite of programs [22]. Basis sets of triple-zeta plus polarization split valence quality (TZVP hereafter) [23] were adopted for all atoms in the complexes. The DFT grid-size was set to standard m3 value.

The computation of the EPR g tensors and hyperfine coupling constants (hcc) for the ^{63}Cu nucleus has been performed using the B3LYP hybrid functional on the geometries optimized at the RI-BP86/TZVP level. In fact, as shown in a previous DFT investigation about copper coordination to prion protein peptides [24], the structural features (bond distances and angles) of Cu-peptide complexes obtained with B3LYP and BP86 are almost identical. However, BP86 calculations are faster, due to the possibility of using the RI technique, and therefore more suited to test a large number of species, such as in the present work. On the other hand, EPR g tensors and hyperfine coupling constants obtained with B3LYP are in better agreement with experimental data, when compared to BP86 results. An extended basis set (14s,10p,5d) [25], augmented by a set of diffuse s, p and d functions (with exponents equal to 0.01, 0.03087, and 0.1, respectively) and contracted to (9s, 7p, 4d) was adopted for the Cu atom. The IGLO-II basis was adopted for all other atoms [26]. EPR

hyperfine coupling constants (hcc) were calculated, explicitly taking Spin-Orbit (SO) contributions into account, with the one-centre and mean field approximation (AMFI) [27] for the two electron terms (see Ref. [28] for a complete discussion of the SO operators). For the calculations of g tensors, the gauge origin has been set to the centre of electronic charge. Calculations of EPR properties have been carried out using the ORCA suite of programs [29].

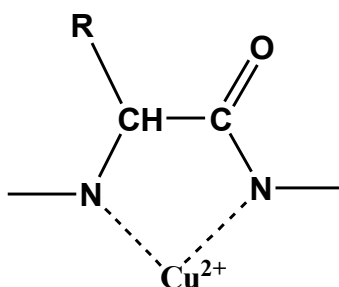
3. Results and Discussion

3.1 Geometries and relative energies of isomers of $[\text{Cu}(\text{Ac-PHREN-NH}_2)(\text{H}_2\text{O})]$ and $[\text{Cu}(\text{Ac-PHRQN-NH}_2)(\text{H}_2\text{O})]^{+1}$ complexes.

According to the experimental evidences, Cu(II) coordination to the Ac-PHREN-NH₂ peptide at pH 8 involves two N⁻ anions from the deprotonate amide groups of the peptide bonds, the carboxylate group of Glu and a nitrogen atom from the His side chain. In addition, a water molecule should also be coordinated to the metal ion [16]. Before discussing which deprotonated N⁻ groups are actually involved in Cu(II) coordination, some considerations are in order. Previous results obtained studying coordination compounds formed by Cu(II) with fragments of the prion protein [24, 30-33], as well as simple considerations about the preferential formation of 5- and 6-member rings when dealing with chelate ligands, lead to the conclusion that the two N⁻ anions involved in Cu(II) coordination must be contiguous. Therefore, only three configurations are possible, namely those involving the **1-2**, **2-3** or **3-4** deprotonated amide groups (see Scheme 1 for amide groups labeling). We preliminary analyzed the conformation of these three configurations at Molecular Mechanics (MM) level, in order to quickly understand which is the most likely Cu²⁺ coordination geometry. The preliminary MM analysis revealed that, when the **1-2** or **3-4** configurations are taken into account, the histidine (in the **1-2** configuration) or the glutammic acid (in the **3-4** configuration) side chains occupy the apical position in the metal coordination sphere. In fact, since the side chains of His (in the **1-2** configuration) or Glu (in the **3-4** configuration) are in between the deprotonated N⁻ groups that form the planar 5-member cycle (Scheme

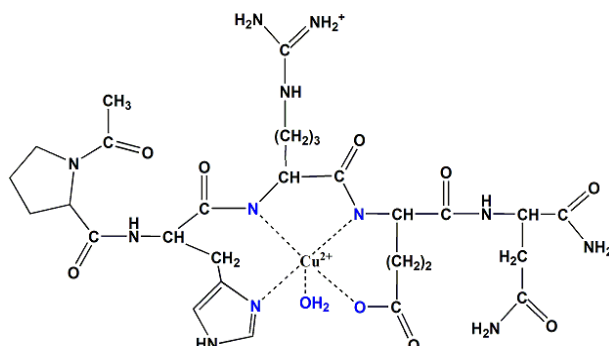
2), the aminoacid side chain must necessarily coordinate Cu^{2+} in apical position due to steric restrictions. As a consequence, in complexes characterized by **1-2** or **3-4** binding mode the water molecule must occupy an equatorial position in the Cu(II) coordination environment.

Scheme 2. Five-member ring formed in Ac-PHREN-NH_2 upon coordination to Cu(II) of the two adjacent deprotonated amide groups.



MM results indicate also that the **1-2** and **3-4** binding modes induce some strain in the peptide chain. In particular, the side chain of the coordinating aminoacid which does not belong to the five member ring (Glu in **1-2** complexes, His in **3-4** structures) is far from the metal center and its coordination to the metal atom is always accompanied by some strain of the peptide backbone. MM analysis of the **2-3** binding mode reveals that the imidazole ring of His and the carboxylate group of Glu can occupy equatorial positions in the Cu(II) coordination sphere without inducing steric strain in the molecule. These results suggest that the structures featuring the **2-3** binding mode most likely correspond to the lowest energy isomers (see Scheme 3).

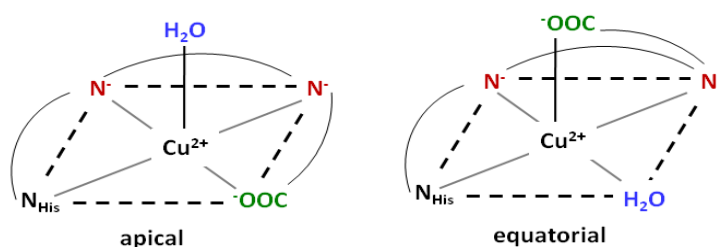
Scheme 3. Schematic representation of the **2-3** binding mode in complexes between Cu(II) , the peptide Ac-PHREN-NH_2 and a water molecule.



Therefore, we have carried out a thorough DFT investigation of the conformational properties of coordination compounds featuring the **2-3** binding mode. However, for the sake of completeness, we have evaluated by DFT also some structures characterized by the **1-2** or **3-4** binding mode, to quantify the strain of the peptide chain.

To select reasonable starting structures for DFT optimization, we have carried out classical MD simulations (see Methods) of the [Cu(Ac-PHREN-NH₂)(H₂O)] complex featuring the **2-3** binding mode, selecting snapshots along the MD trajectory which correspond to low energy conformations. The protocol adopted to explore the configurational space of the system is the following. The structures characterized by the **2-3** binding mode can be ideally divided into two families: the first containing structures featuring an apical water molecule (and consequently an equatorial Glu side chain), and the second containing structures featuring an equatorial water molecule (Scheme 4).

Scheme 4. The two families of [Cu(Ac-PHREN-NH₂)(H₂O)] complexes characterized by the **2-3** binding mode. Apical and equatorial labels are referred to the position of the water molecule in the Cu²⁺ coordination sphere.

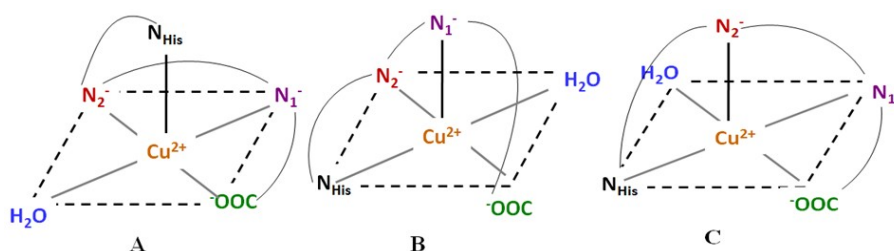


We initially generated “ideal” (dihedral angle of the base of the pyramid equal to zero, molecular axes of the apical ligand exactly perpendicular to the base) structures characterized by the N_3O_2 coordination environment and featuring an apical or equatorial water molecule. The value of Cu-N⁻ distances has been set to 2.000 Å, according to the BP86 optimized Cu-N distances in the [Cu(HGGG)] system [33]. For the Cu-N_{His} and Cu-O_{Glu} distances, we adopted the value of 1.894 and 2.007 Å obtained respectively from BP86/TZVP gas-phase optimization of the Cu-His and Cu-Glu adducts. Regarding the Cu-O distance to describe Cu-H₂O binding, we considered as experimental reference the XRD value of 2.4 Å in *cis*-Cu(Gly)₂·H₂O, which fits well also with the analysis of Sobolović et al. [34], which measured a mean value of 2.4±0.1 Å for Cu-O distances in 10 aminoacid/water Cu(II) complexes.

Using the two “ideal” structures as templates, we generated apical and equatorial conformations from molecular dynamics simulations in which the Cu(II) coordination distances were constrained to the above mentioned reference values. Then, for each structure obtained, we first optimized its geometry at the MM level of theory, and the corresponding minimum energy structure was used as starting point of a two-step DFT optimization. In the first step the Cu(II) coordination was constrained, allowing only the peptide moiety to relax. In the second step the whole geometry was optimized removing the constraints.

The relevant optimized geometry parameters, as well as computed relative energies of all structures characterized by the **2-3** binding mode are collected in Table 1. It must be noted that in structures characterized by an apical water molecule the N_3O_2 coordination can be realized only in one way, whereas in the case of equatorial water binding the isomer shown in Scheme 4 is not the only plausible possibility, but other three ligands arrangements, referred to as the A, B and C, can in principle exist (Scheme 5).

Scheme 5. The three subfamilies of [Cu(Ac-PHREN-NH₂)(H₂O)] complexes characterized by the **2-3** binding mode and the water molecule in equatorial position.



Before discussing the fine details of the various structures optimized, it is interesting to highlight some trends obtained by our DFT investigation. Except for a few structures, geometry optimizations usually led to tetrahedral distortion of the initial square pyramidal coordination. In addition, the H₂O ligand often resulted loosely bound or non-bound to Cu(II), yielding 4-coordinated N₃O structures. It is also worth noting that when starting from structures featuring an equatorial water ligand, the H₂O molecule always leaved the Cu(II) coordination sphere, and it was replaced in the equatorial position by the oxygen atom of the glutammic acid residue. Therefore, structures characterized by equatorial coordination of the H₂O molecule should not correspond to stable forms of the [Cu(Ac-PHREN-NH₂)(H₂O)] complex.

In the exploration of the conformation space of [Cu(Ac-PHREN-NH₂)(H₂O)], we often found that structures differ only for the number of intramolecular H-bonds. As shown in the following, when a complex is explicitly solvated by a reasonable number of H₂O molecules, intramolecular interactions are often lost because the side-chains of the residues not involved in Cu(II) coordination (in particular arginine) prefer to form H-bonds with the solvent. Nevertheless, even if we are aware of the limits of such approach, we carried out most DFT calculations neglecting solvent effects. In fact, the explicit description of a shell of water molecules surrounding the [Cu(Ac-PHREN-NH₂)(H₂O)] complex would be computationally too demanding. Indeed, since one of the main purposes of this study was the elucidation of the position of the H₂O ligand in the Cu(II) coordination sphere, in principle solvent effects due to bulk water could be described using an implicit solvent model. However, it was already observed that DFT tends to underestimate H₂O-Cu(II) binding in similar complexes [33], and we noted that the adoption of an implicit solvent model makes the H₂O-Cu(II) bond even weaker (vide

infra), not allowing to properly study H₂O coordination to Cu(II) in this class of compounds.

The lowest energy structure characterized by the **2-3** binding mode features an apical water molecule (**1**; Figure 1). In **1** the N₃O₂ square pyramidal Cu(II) coordination environment is characterized by a dihedral angle $\hat{\curvearrowright}$ (defined by the atoms N_{HIS}-N₁⁻-N₂⁻-O⁻ forming the base of the square pyramid) equal to 0.4 degree, and a Cu-O_{H₂O} distance = 2.645 Å. Other five structures are characterized by a Cu-O_{H₂O} distance lower than 3.0 Å (**8**, **11**, **12**, **14**, **19**; Table 1). When compared to **1**, these structures are characterized by a smaller number of H-bonds, except for structure **14**, which however is characterized by a loosely bound H₂O molecule (Cu-O_{H₂O} = 2.753Å). The structure with the shortest Cu-O_{H₂O} distance (**12**) features an apical water molecule and is 13.7 kcal·mol⁻¹ higher in energy than **1**. In this context it is also worth noting that **1** and **12** differ for the number of intramolecular H-bonds (4 in **12**, 6 in **1**; d(H···O) ≤ 2.0Å), underlying the role played by this type of interactions for the relative energies of the structures considered.

The electronic spectra of Ac-PHREN-NH₂ and the lowest energy [Cu(Ac-PHREN-NH₂)(H₂O)] form (structure **1**) have been computed at BP86/TZVP TDDFT level and are shown in Figure 2. The first 100 transitions have been computed and excitation energies, corresponding oscillation strengths and MO compositions for the first 30 transitions are collected in Table 2. It must be noted that the BP86 functional is not the optimal choice to reproduce excitation energies, in particular when metal-to-ligand charge transfer (MLCT) bands are considered, but the large size of the systems investigated did not allow us to adopt more suited functional, as PBE0 [35]. In any case, it can be concluded that computed UV-Vis bands of Ac-PHREN-NH₂ are mainly due to the histidine residue, and correspond to two intense absorptions at 266 nm and 240 nm, and one weak band at 320 nm. In the corresponding Cu(II) complex the two intense bands are red-shifted by about 100 nm. The bands in the range 1000-500 nm (Table 2) are of SOMO-n(β)→SUMO(β) type (n=0-14, Singly Occupied and Singly Unoccupied MO, respectively) and essentially correspond to the Cu(II) d-d transitions. In particular, the first five transitions that involve the d-Cu(II) MOs are of d-d type. The most intense band is at 689 nm (excitation 6), in fairly good agreement with the 640 nm absorption reported by La Mendola et al. [16]h

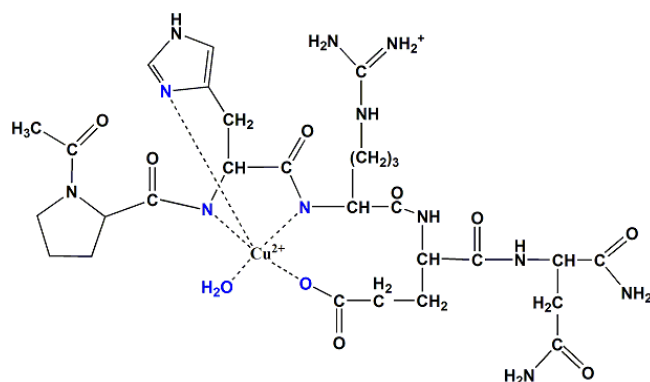
As already pointed out, **2-3** isomers characterized by equatorial disposition of the water molecule do not correspond to minima on the potential energy surface, suggesting an intrinsic instability of this coordination mode. In order to quantify the destabilization of structures featuring an equatorially coordinated water molecule, we have also optimized a structure (**20**) which is obtained from **1** after exchange of the positions of the water molecule and the carboxylate group. In order to maintain the Cu(II) coordination environment, the position of the two coordinated oxygen atoms from the water molecule and the carboxylate group have been constrained to the Cu-O distances observed in **1**. Although this evaluation is qualitative, we found that structure **20** is more than 30 kcal·mol⁻¹ higher in energy than **1**.

In order to evaluate if an extra water molecule might be bound to Cu(II), leading to a pseudo octahedral coordination environment, we further investigated complex **1** after addition of a second H₂O molecule in apical position. However, the optimization of such a species led to a square pyramidal structure in which the extra water molecule moved away from the metal atom (final Cu-O_{H₂O} distance = 3.567Å). Going back to the issue of the computational approach adopted in this study to describe solvation effects, we also optimized the lowest energy structure (**1**) including solvent effects through the implicit solvent model COSMO (COnductor-like Screening MOdel) [36-38]. The main effect observed in COSMO calculations was the removal of the H₂O molecule from the Cu(II) coordination sphere, which consequently becomes tetrahedrally distorted (δ moves from 0.4 to 30.7 degree).

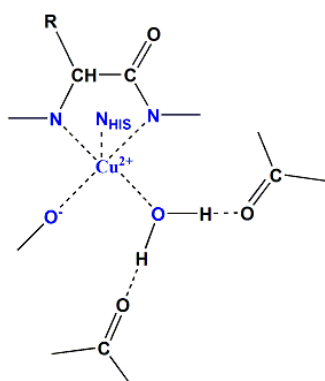
Finally, we have also evaluated the lowest energy complex **1** when explicitly solvated (**1W**). To address this point, structure **1** was solvated with a cluster of 55 H₂O molecules, whose position was optimized by mean of a short MD simulation (1500 ps) and successive MM geometry optimization. The solvated structure obtained in this way was successively optimized at DFT level using the same level of theory adopted for the non-solvated models (BP86/TZVP). Upon geometry optimization, the Cu(II) coordination does not change significantly. The H₂O molecule initially at 2.645Å gets closer to the metal center (2.445Å) and the δ dihedral angle increases of only 2 degree. Notably, no intramolecular H-bonds are found because the aminoacid sidechains are all involved in H-bond with the solvent.

DFT results obtained studying structures characterized by the **1-2** and **3-4** binding mode indicate that these isomers are higher in energy with respect to the lowest energy **2-3** isomer (**1**). In particular, most of the **1-2** structures (**21**, **22**, **23**, **26** and **28**; Table 1) feature a Cu(II) coordination with the histidine residue in apical position (see Scheme 6) and are higher in energy with respect to isomer **1** by at least 13.7 kcal·mol⁻¹.

Scheme 6. Schematic representation of the **1-2** binding mode in complexes between Cu(II), the peptide Ac-PHREN-NH₂ and a water molecule.



Scheme 7. Schematic representation of the Cu(II) coordination environment in the lowest energy complex characterized by the **1-2** binding mode (structure **21**).

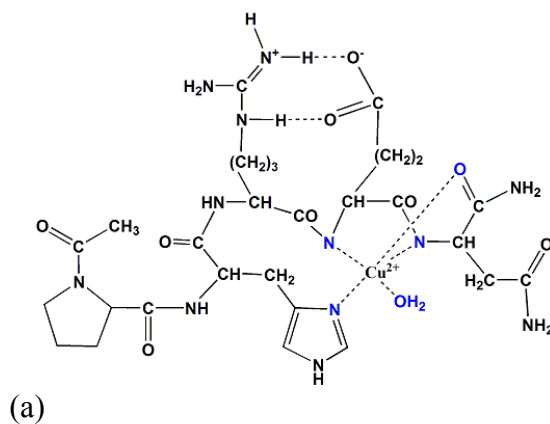


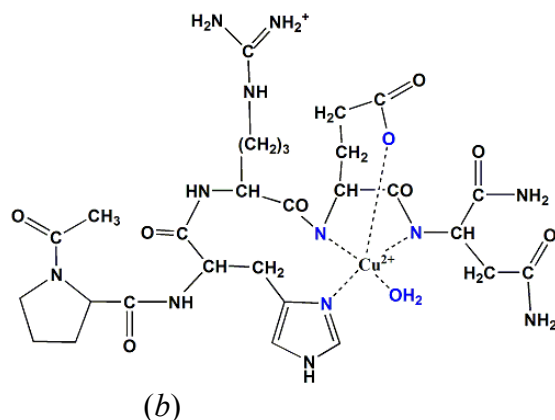
The three lowest energy **1-2** structures feature square pyramid coordination with a small tetrahedral distortion. The lowest energy **1-2** form (structure **21**) is shown in Figure 3. It is interesting to point out that the H₂O ligand is coordinated to Cu(II) with a much shorter Cu-O distance (2.031Å) compared with the **2-3** structures. In fact, in **21** the Cu(II)-H₂O binding is “assisted” by the formation of two H-bonds formed with the two CO groups of the Glu sidechain and the peptide bond between Pro and His (Scheme 7).

During the conformational sampling of structures characterized by the **1-2** binding mode we have found three 4-coordinated structures (**24**, **25** and **27**), in which the histidine sidechain is not coordinated to Cu(II). However, these isomers are at least 8.6 kcal·mol⁻¹ higher in energy than the lowest energy **1-2** form (**21**).

The analysis of the MD simulations of the **3-4** configuration revealed a larger number of possible copper coordinations. In particular, several structures obtained from the MD simulations can be classified in two families (Scheme 8).

Scheme 8. Schematic representation of the two families of complexes between Cu(II), the peptide Ac-PHREN-NH₂ and a water molecule, characterized by the **3-4** binding mode.





In structures **29**, **31**, **36** and **40** (Scheme 8; family *a*) the carboxylate of the glutamic acid forms an electrostatic interaction with the side-chain of arginine, and consequently the carbonyl group of the C-terminal amide group enters the Cu(II) coordination sphere in equatorial position. Among the family *a* species, structure **29** is the lowest in energy (+8.1 kcal·mol⁻¹ with respect to **1**).

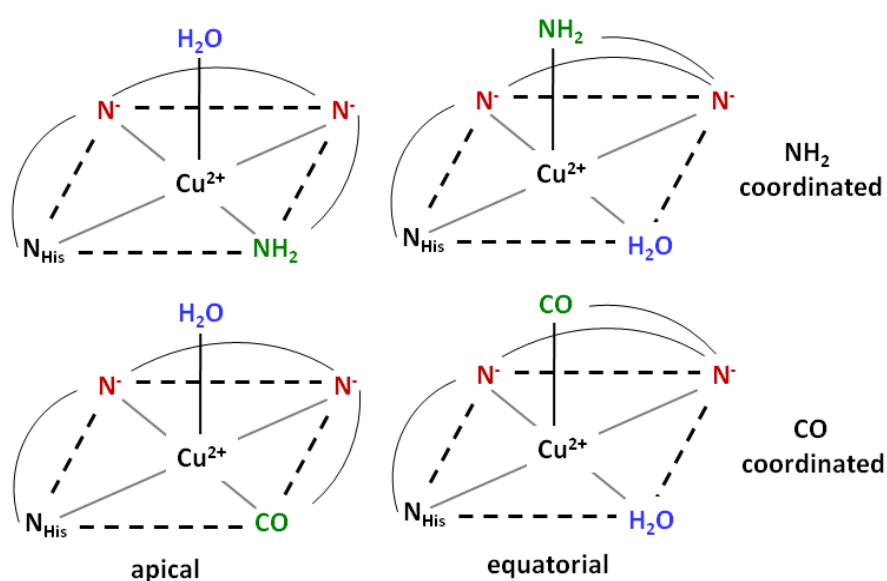
In structures **32**, **35**, **38**, **41** and **43** (Scheme 8; family *b*) the carboxylate group of Glu is coordinated to Cu(II) in apical position, and therefore the water molecule occupy the equatorial position. All these structures are genuine equatorial forms, but, however, are even higher in energy than the constrained **2-3** equatorial form (structure **20**).

The other **3-4** isomers obtained from the conformational sampling are higher in energy and feature a different Cu(II) coordination environment. It is worth noting that the 3-coordinated structure **30** (2N⁻ and N_{His}) lies only 1.6 kcal·mol⁻¹ higher in energy with respect to the lowest energy **3-4** structure (**29**) because it is stabilized by the formation of 6 intramolecular H-bonds. Structures **33** and **34**, which are characterized by substitution of the coordinated H₂O with a CO group of the peptide chain, lie at least 29.6 kcal·mol⁻¹ higher in energy with respect to **1**. In **34** also the histidine sidechain leaves the Cu(II) coordination sphere, giving a 4-coordinated form. Structure **37** is the only **3-4** form that features the histidine sidechain in apical position, similarly to the **1-2** forms.

In order to better understand the role of the glutamate side chain in Cu(II) binding, and prompted by the previous observation that

cleavage of the E67-N68 peptide bond compromises the angiogenic activity of Ang [39], La Mendola et al. investigated also Cu(II) binding to the E→Q variant of the Ac-PHREN-NH₂ peptide ($[\text{Cu}(\text{Ac-PHRQN-NH}_2)(\text{H}_2\text{O})]^{+1}$) observing at pH 8 a value of the A_{\parallel} component of the ⁶³Cu hyperfine coupling constant slightly larger than that measured for the wild type species [16].

Scheme 9. The four conformation subfamilies of the $[\text{Cu}(\text{Ac-PHRQN-NH}_2)(\text{H}_2\text{O})]^{+1}$ complex. The labels apical and equatorial are referred to the position of the water molecule in the Cu²⁺ coordination sphere.



In principle, four different types of coordination can be observed in complexes between Cu(II) and Ac-PHRQN-NH₂ (Scheme 9), in which the glutamine residue can be coordinate to Cu(II) via the NH₂ or CO group of its side-chain, and the H₂O ligand can occupy apical or equatorial position. As well know from experimental evidences at neutral pH the amidic NH₂ group should not coordinate the Cu(II) ion. However, to test the computational approach employed in this investigation it is interesting also to consider this type of coordination. Sampling of the configuration space of the $[\text{Cu}(\text{Ac-PHRQN-NH}_2)(\text{H}_2\text{O})]^{+1}$ complex has been carried out using the same approach

adopted for the wild-type peptide. In particular, starting from the lowest energy forms of the Ac-PHREN-NH₂ complex, we introduced the E→Q substitution and successively carried out MD simulations to obtain starting point structures for DFT optimizations. We extracted from the MD trajectories 8 isomers characterized by apical (4 structures) or equatorial (4 structures) H₂O coordination, and for each structure we evaluated both CO and NH₂ coordination of the glutamine side-chain, obtaining a total number of 16 starting point structures for DFT optimization.

In general, in structures characterized by coordination of the NH₂ group of Gln, the Cu-N distance is about 2.3 Å and the NH₂ group is in equatorial position. The corresponding forms in which the CO amidic group is coordinated in equatorial position to Cu²⁺ features a Cu-O distance within 2.1-2.2 Å.

Relative energies of the various DFT optimized structures, as well as relevant bond distances for [Cu(Ac-PHRQN-NH₂)(H₂O)]⁺¹ complexes are collected in Table 3. Notably, all [Cu(Ac-PHRQN-NH₂)(H₂O)]⁺¹ low energy structures are characterized by a significant tetrahedral distortion. The lowest energy [Cu(Ac-PHRQN-NH₂)(H₂O)]⁺¹ isomer (**Q1**, see Figure 4) features the H₂O molecule in equatorial position, whereas the carbonyl group of the glutamine sidechain is coordinated to the Cu(II) atom in apical position. The high stability of **Q1** may be due to the large number of intramolecular H-bond (8). In fact, the isomer **Q2**, which is characterized by a coordination environment similar to that of **Q1**, but with a lower number of H-bond (5), is only 1.3 kcal mol⁻¹ higher in energy than **Q1**. **Q3** is also close in energy to **Q1** (ΔE = 2.2 kcal·mol⁻¹). The latter isomer is a four-coordinate species in which the glutamine sidechain has left the Cu²⁺ coordination sphere and the equatorial H₂O molecule is coordinated to the Cu atom with the shortest O-Cu distance (2.12 Å). Similarly, the isomer **Q5** is also 4-coordinated, but it is 10.3 kcal·mol⁻¹ higher in energy, possibly due to a smaller number of intramolecular H-bond interactions.

The lowest-energy isomer in which the carbonyl group of the glutamine sidechain is coordinated to the Cu(II) atom in the equatorial position is **Q7**, which is 13.2 kcal·mol⁻¹ higher in energy than **Q1**. Interestingly, **Q7**, as well as all other isomers with the glutamine carbonyl group coordinated at the equatorial position (**Q15-Q18**), are 4-coordinated species, because the H₂O molecule leaves the Cu(II) coordination sphere.

The structures **Q8** and **Q9** are similar and are the lowest in energy among those in which the NH₂ group of glutamine is coordinated to Cu(II) atom. However, their energy is 15.2 kcal·mol⁻¹ higher than in **Q1**. These two isomers are also 3.0 kcal·mol⁻¹ less stable than **Q7** even if two more H-bond interactions are present. **Q8** and **Q9** are penta-coordinated species with the H₂O molecule weakly coordinated in apical position at about 2.5 Å from the Cu(II) atom. On the other hand, the other two isomers in which the NH₂ group of glutamine is coordinated to Cu(II) (**Q13** and **Q14**) are fully tetra-coordinated species.

3.2 EPR hyperfine coupling constants and g values of isomers of [Cu(Ac-PHREN-NH₂)(H₂O)] and [Cu(Ac-PHRQN-NH₂)(H₂O)]⁺¹ complexes.

In the previous section it was pointed out that the energy criterion is not sufficient to confidently predict the coordination mode of Cu(II) in the [Cu(Ac-PHREN-NH₂)(H₂O)] and [Cu(Ac-PHRQN-NH₂)(H₂O)]⁺¹ complexes, because of the different number and strength of intramolecular H-bonds in the isolated isomers, which affect their energy ranking. Therefore, in order to get a better insight into the coordination mode of these complexes, EPR ⁶³Cu hyperfine coupling constants (⁶³Cu-hcc) and g-tensors of the most relevant isomers have been calculated, and compared with the experimental A_{||} and g_{||} values reported in Ref. [16]. Results are collected in Tables 4 and 5. As well known, for the Cu²⁺ ion the spin-orbit coupling (relativistic) contributions are important, due to its relatively large nuclear charge. Therefore, the Fermi contact (A_{FC}) and the principal values of the dipolar tensor (A_{dip}) are discussed in connection with relativistic corrections (Pseudo Contact (A_{PC}), and A_{dip2}).

All of the complexes investigated, with the exception of **5** and **10** are characterized by a roughly planar N₃O disposition of ligands ($\hat{\angle} < 20^\circ$; see Table 1). ⁶³Cu-hcc tensor components of these complexes feature an axial symmetry, as deduced from the observation that the A_{xx} (A_{xx} ≡ A_{iso} + A₁₁) and A_{yy} (A_{yy} ≡ A_{iso} + A₂₂) elements of tensor do not differ by more than 20 MHz. However, for these complexes the A_{||} (A_{||} ≡ A_{iso} + A₃₃) component is significantly overestimated with respect to the

experimental value (see Table 4). In this respect, it is worth noting that the A_{\parallel} component appears to be suppressed by the axial coordination of the H_2O molecule, as deduced by the observation that in the 4-coordinated complexes ($Cu-O_{H_2O} > 3 \text{ \AA}$) $A_{\parallel} = 540\text{-}556 \text{ MHz}$, whereas in **8** and **11** ($Cu-O_{H_2O} \approx 2.49 \text{ \AA}$) $A_{\parallel} = 507\text{-}517 \text{ MHz}$. For the latter complexes the A_{\parallel} value is in better agreement with the experiment. Therefore, the comparison between experimental and computational EPR parameters might suggest a N_3O_2 (slightly distorted) square pyramidal Cu(II) coordination environment, with the H_2O molecule weakly coordinated in axial position.

Isomers **5** and **10**, which are 5.5 and $12.2 \text{ kcal}\cdot\text{mol}^{-1}$, respectively, higher in energy than **1**, are characterized by a significant tetrahedral distortion in the N_3O arrangement of ligands. For these two isomers the A_{\parallel} component is in very good agreement with the experimental value. In fact, the calculated tensor of **5** slightly deviates from the axial symmetry ($\hat{A}_{\text{dip}} = 34$) by a value which may not be resolved in the experimental spectrum. A larger deviation ($\hat{A}_{\text{dip}} = 72$) is observed for the complex **10**. It is interesting to note that **5** and **10** are both characterized by one of the two N-Cu distances significantly shorter than in the other complexes, a feature which could also account for the depression of the A_{\parallel} component when moving from the square planar to the distorted configuration.

The A_{\parallel} value computed for isomer **20**, which features an equatorial H_2O molecule constrained to 2.407 \AA from the Cu(II) atom, is significantly lower than the A_{\parallel} value obtained in isomers characterized by axial coordination of the H_2O molecule, and apparently in fairly good agreement with the experimental value. However, equatorial coordination of the H_2O molecule can be ruled out since the calculated tensor components significantly deviate from ideal axial symmetry (see \hat{A}_{dip} in Table 4), a feature which is not compatible with the experimental EPR spectrum.

The computed values of g components are only slightly affected by a particular coordination mode, as reported for other copper-peptide complexes [33], and are, therefore, of little help in discriminating different isomeric forms. In general, we observe for the calculated g-tensors that the trend is similar to that discussed for the $^{63}\text{Cu}\text{-hcc}$ parameters (see Table 5). The g-tensor computed for isomer **20** exhibits a significant rhombic distortion, whereas all other isomers are characterized by an almost ideal axial symmetry of the g-tensor. In

addition, the g_{\parallel} component of the penta-coordinated complexes **8** and **11**, is slightly larger than that of the tetra-coordinated complexes, and in closer agreement with the experimental value.

The experimental A_{\parallel} and g_{\parallel} values obtained studying $[\text{Cu}(\text{Ac-PHRQN-NH}_2)(\text{H}_2\text{O})]^{+1}$ differ from those measured for the $[\text{Cu}(\text{Ac-PHREN-NH}_2)(\text{H}_2\text{O})]$ complex, leading to the suggestion that the carboxylate group is involved in Cu(II) coordination in the latter [16]. EPR parameters of selected $[\text{Cu}(\text{Ac-PHRQN-NH}_2)(\text{H}_2\text{O})]^{+1}$ complexes are reported in Tables 6 and 7. As a general remark, we observe for most of the species investigated that, contrary to the experimental data, the A_{\parallel} component of the ^{63}Cu -hcc tensor is lower than that calculated for the isomers of the $[\text{Cu}(\text{Ac-PHREN-NH}_2)(\text{H}_2\text{O})]$ complex. Inspection of the ^{63}Cu -hcc tensor contributions reported in Table 6 shows that this behavior is mainly due to the isotropic term. The lowest energy isomers **Q1** and **Q2** feature a similar distorted square pyramidal arrangement of ligands, with the $\text{O}_{\text{H}_2\text{O}}$ and the O_{Gln} atoms coordinated in basal and apical positions, respectively. The main difference in the coordination sphere between the two isomers lies in the $\text{O}_{\text{H}_2\text{O}}$ -Cu and O_{Gln} -Cu distances, which in **Q1** are about 0.1 and 0.04 Å shorter than in **Q2**. The A_{\parallel} component of the ^{63}Cu -hcc tensor of **Q1** and **Q2** is smaller than the experimental value by about 90 and 40 MHz, respectively. In addition, for both isomer, the tensor is characterized by a large rhombic distortion, a feature which is not compatible with the experimental spectrum. **Q3** and **Q5** are 4-coordinated isomers in which the N_3O coordination sphere is completed by the H_2O molecule. The **Q3** complex, which is roughly planar and only 2 kcal mol⁻¹ higher in energy than **Q1**, is characterized by a ^{63}Cu -hcc tensor with axial symmetry. The A_{\parallel} component of the calculated tensor is about 60 MHz larger than the experimental value. In addition, the calculated A_{\parallel} value is very similar to that of the square planar species in the $[\text{Cu}(\text{Ac-PHREN-NH}_2)(\text{H}_2\text{O})]$ complex, even if the carboxylate ligand in the equatorial position is replaced by the H_2O molecule. The arrangement of ligands in **Q5**, differently from **Q3**, is characterized by a significant tetrahedral distortion. Notably, the A_{\parallel} component of the ^{63}Cu -hcc tensor is smaller than in **Q3**, and matches within 10 MHz the experimental value. In addition, the tensor deviates only slightly from the ideal axial symmetry ($\hat{A}_{\text{dip}} = 34$). Furthermore, this is the only isomer for which the replacement of the carboxylate ligand with the H_2O molecule increases the A_{\parallel} value by 30

Mhz, a value in very good agreement with that observed in the experiment when replacing the E residue with Q. **Q4** and **Q6** are 5-coordinated species in which the arrangement of the ligands approaches the trigonal bipyramidal symmetry. The calculated hcc tensor for these two tensors is both quantitatively and qualitatively very different from the experimental one, further indicating that the structure of the complex in solution should not deviate significantly from the square planar or square pyramidal symmetry. **Q8** is the lowest energy isomer characterized by the equatorial coordination of four nitrogen atoms and the axial coordination of the H₂O molecule. The ⁶³Cu-hcc tensor of this isomer has axial symmetry, and the A_{||} value is about 40 MHz lower than the experimental one, indicating that this species cannot be excluded only on the basis of the computed EPR parameters. Finally, we have also calculated the ⁶³Cu-hcc tensor of **Q7**, **Q11** and **Q16** in which the carbonyl group of the glutamine sidechain is coordinated to the Cu(II) atom in the equatorial position, and featuring different tetrahedral distortions ($\hat{\alpha} = 37.5^\circ, 16.3^\circ$ and 5.7° for **Q7**, **Q11** and **Q16**, respectively). In **Q11** the H₂O molecule is also weakly coordinated to the apical position. As shown in Table 6, the A_{||} component in this latter species is underestimated by more than 100 MHz with respect to the experimental value, and the tensor is characterized by a large rhombic distortion. In **Q7** the A_{||} component is about 50 MHz smaller than in the experiment and the tensor is less affected by the rhombic distortion. In **Q16**, the A_{||} component of the ⁶³Cu-hcc tensor is in very good agreement with the experiment, and the tensor further approaches the axial symmetry. However, **Q16** is 37.4 kcal·mol⁻¹ higher in energy than **Q1**, an energy difference which allow to exclude that such isomer is present in aqueous solution. As observed for the corresponding ⁶³Cu-hcc parameters, the g tensors computed for **Q1**, **Q2**, **Q4** and **Q11** are characterized by a significant rhombic distortion, whereas in **Q3**, **Q8** and **Q11** the tensors approach to the axial symmetry. It is worth noting that the g_{||} component of **Q16**, which features the glutamine carbonyl ligand in the equatorial position, is smaller than that of the species with the H₂O molecule in the equatorial position, a trend opposite with respect to the experimental results.

4. Conclusions

In this paper we have reported results obtained studying Cu^{2+} coordination to the Ac-PHREN-NH₂ peptide, that represents the Cu^{2+} binding portion of angiogenin, a protein playing a key role in the process of angiogenesis. Using as a reference the recent experimental results reported by La Mendola et al. [16], we have used classical MM/MD calculations followed by DFT optimizations to explore the configurational space of the $[\text{Cu}(\text{Ac-PHREN-NH}_2)(\text{H}_2\text{O})]$ complex.

EPR experiments suggested that Cu(II) is coordinated in a N₃O square planar environment characterized by a significant tetrahedral distortion, where the Cu(II) coordination sphere is composed by 2N⁻ of deprotonated peptide bonds, the nitrogen atom of the histidine group and the carboxylate group of the glutamic residue. An alternative N₃O₂ (less distorted) square pyramidal coordination in which the oxygen atom of a weakly coordinated water molecule occupies the apical position cannot be excluded. In addition a detailed comparison of computed minimum energy structures and EPR parameters suggests that the binding of Cu(II) to the deprotonated amide groups of the peptide bonds takes place between the residues H-R and R-E.

We have also investigated the configuration space of the E→Q mutated system $[\text{Cu}(\text{Ac-PHRQN-NH}_2)(\text{H}_2\text{O})]^{+1}$. In this case, computational results led to the conclusion that, similarly to the $[\text{Cu}(\text{Ac-PHREN-NH}_2)(\text{H}_2\text{O})]$ complex, the coordination mode is described by a N₃O square planar environment with a large tetrahedral distortion, in which the carboxylate group is replaced by a H₂O molecule. Finally, the structures which better fit the experimental EPR parameters does not correspond to the most stable isomers, further indicating that the energy ranking of isomers, derived in the absence of solvent water molecules, is not a reliable criterion in the study of speciation of metal-peptide complexes in aqueous solution.

Acknowledgments

This work was supported by the PRIN Project N 200875WHMR.

References

1. Folkman J (1989) *J Natl Cancer Inst* 82:4
2. Bussolino F, Mantovani A, Persico G (1997) *Trends Biochem Sci* 22:251
3. Kishimoto K, Liu S, Tsuji T, Olson K A, Hu G (2005) *Oncogene* 24:445
4. Fett JW, Strydom DJ, Lobb RR, Alderman EM, Bethune JL, Riordan JF, Vallee BL (1985) *Biochemistry*, 24:5486
5. Gao X, Xu Z (2008) *Acta Biochim Biophys Sin* 619
6. Badet J, Soncin J, Guitton JD, Lamare O, Cartwright T, Barritault D (1989) *Proc Natl Acad Sci USA*, 86:8427
7. McAuslan BR, Reilly W (1980) *Exp Cell Res*, 130:147
8. Hu G-F, (1998) *J Cell Biochem*, 69:326
9. Lequin O, Thuring H, Robin M, Lallemand J-Y (1997) *Eur J Biochem*, 250:712
10. Millhauser GL (2007) *Annu Rev Phys Chem*, 58:299
11. Solioz M, Vulpe C (1996) *Trends Biochem Sci*, 21:237
12. Soncin F, Guitton JD, Cartwright T, Badet J (1997) *Biochem Biophys Res Commun* 236:604
13. Joyce BK, Cohn M (1969) *J Biol Chem* 244:811
14. Acharya KR, Shapiro R, Allen SC, Riordan JF, Vallee BL (1994) *Proc Natl Acad Sci USA* 91:2915
15. Smyth DG, Stein WH, Moore S (1963) *J Biol Chem* 238:227
16. La Mendola D, Magri A, Vagliasindi LI, Hansson O, Bonomo RP, Rizzarelli E (2010) *Dalton Trans*, 39:10678
17. Halgren TA (1996) *J Comput Chem* 17:490
18. Halgren TA (1999) *J Comput Chem* 20:720
19. Becke AD (1988) *Phys Rev A*, 38:3098
20. Perdew JP (1986) *Phys Rev B*, 33:8822
21. Eichkorn K, Weigend F, Treutler O, Ahlrichs R (1997) *Theor Chem Acc* 97:119
22. Ahlrichs R, Bar M, Haser M, Horn H, Kolmel C (1989) *Chem Phys Lett* 62:165
23. Schafer A, Huber C, Ahlrichs R (1994) *J Chem Phys*, 100:5829
24. Bruschi M, De Gioia L, Mitric R, Bonacic-Koutecky V, Fantucci P (2008) *Phys Chem Chem Phys*, 10:4573
25. Schafer A, Horn H, Ahlrichs R (1992) *J Chem Phys* 97:2571
26. Kutzelnigg W, Fleischer U, Schindler M, (1990) in 'The IGLO-Method: Ab Initio Calculation and Interpretation of NMR Chemical Shifts and Magnetic Susceptibilities', Springer-Verlag
27. Hess BA, Marian CM, Walhgren U, Gropen O (1996) *Chem Phys Lett* 251: 365
28. Kaupp M, Buhl M, Malkin VG (2004) in 'Calculation of NMR and EPR Paramters: Theory and Applications', Wiley-VCH
29. Neese F (2003) *J Chem Phys* 117:3939
30. Marino T, Russo N, Toscano M (2007) *J Phys Chem B*, 111:635

31. Marino T, Russo N, Toscano M (2011) *Int J Quantum Chem*, 111:1152
32. Pushie MJ, Rauk A (2003) *J Biol Inorg Chem*, 8:53
33. Franzini E, De Gioia L, Fantucci P, Zampella G, Bonacic-Koutecky V (2003) *Inorg Chem Commun*, 6:650
34. Sabolovic J, Tautermann CS, Loerting T, Liedl KR (2003) *Inorg Chem*, 42:2268
35. Adamo C, Scuseria G, Barone V, (1999) *J Chem Phys* 111:2889
36. Klamt A (1995) *J Phys Chem A*, 99:2224
37. Klamt A (1996) *J Phys Chem A*, 100:3349
38. Klamt A, Schüürmann G (1993) *J Chem Soc Perkin Trans*, 2:799
39. Hallahan TW (1991) *Proc Natl Acad Sci USA*, 85:5061

Table 1. Selected geometrical parameters (in Å and degree) and energy differences relative to the lowest energy form (in kcal·mol⁻¹) of the [Cu(Ac-PHREN-NH₂)(H₂O)] isomers. In the table are reported the five Cu-L distances, the dihedral angle that defines the base of the square pyramid of the N₃O₂ coordination sphere and the number of hydrogen bonds (H···O distance less than 2.0Å). CN stands for coordination number.

Isomer	CN	ΔE	Distances					∠ dihedra 1	H- bonds
			Cu- O _{H2O}	Cu-O _{CO}	Cu-N _i	Cu-N _z	Cu-N _{His}		
2-3 binding mode									
1	5	0.0	2.645	2.039	2.016	2.019	2.039	0.4	6
2	4	2.0	3.777	2.001	2.017	2.001	2.000	16.6	5
3	4	2.1	3.863	2.012	2.020	2.000	2.006	12.4	5
4	4	2.1	3.796	2.010	2.012	1.999	1.994	13.5	5
5	4	5.5	3.565	2.001	1.996	1.948	2.016	38.5	2
6	4	6.0	3.374	1.973	1.965	2.034	2.003	9.2	4
7	4	6.2	3.354	1.999	2.029	2.033	2.032	17.6	4
8	5	6.3	2.488	2.014	2.017	2.056	2.029	0.9	3
9	4	8.4	3.684	1.983	1.986	2.022	2.012	8.7	4
10	4	12.2	3.655	1.982	1.932	2.074	1.986	42.1	8
11	5	12.7	2.492	2.020	1.994	2.070	2.028	4.3	5
12	5	13.7	2.348	2.044	2.037	2.023	1.971	3.4	4
13	4	14.0	3.390	2.002	1.963	1.929	2.100	37.3	3
14	5	14.8	2.753	2.009	2.032	1.993	2.047	1.3	6
15	4	15.3	3.912	1.954	1.954	2.016	2.021	35.3	5
16	4	15.5	3.755	2.010	1.990	2.019	2.011	18.2	5

17	4	16.0	3.924	1.962	1.949	2.023	2.020	34.9	5
18	4	17.7	4.073	1.932	1.972	2.002	2.038	35.2	6
19	5	24.6	2.889	1.997	2.000	2.013	2.054	19.6	4
20	5	32.3	2.407	1.997	2.000	2.004	1.994	0.4	6

Table 1. continued

Isomer	CN	ΔE	Distances					$\widehat{\text{dihedra}}_1$	H-bonds
			Cu-O _{H2O}	Cu-O _{CO}	Cu-N ₁ ⁻	Cu-N ₂ ⁻	Cu-N _{HIS}		
2-3 binding mode									
21	5	13.7	2.031	2.042	1.984	2.150	2.281	11.0	8
22	5	13.8	2.023	2.042	1.982	2.130	2.326	14.4	8
23	5	20.0	2.020	2.050	2.050	2.004	2.302	9.6	4
24	4	22.3	1.994	2.041	1.977	1.976		44.0	5
25	4	25.4	1.979	2.070	1.987	1.968		47.2	6
26	5	26.7	2.014	2.112	2.029	2.038	2.237	38.9	6
27	4	26.8	1.973	1.998	1.998	1.985		30.6	3
28	5	31.2	2.017	1.989	2.107	2.063	2.084	47.5	4
3-4 binding mode									
29	5	8.1	2.450	2.242 [*]	1.931	1.999	2.050	31.0	5
30	3	9.7	3.075		1.915	2.023	1.978		6
31	5	16.4	2.544	2.178 [*]	2.012	1.927	2.024	19.7	5
32	5	26.1	2.170	2.010	2.000	2.078	2.168	47.5	5
33	5	26.2	3.414	2.574 2.711 [*]	1.935	1.986	2.003	53.7	6
34	5	35.2	3.641	2.085 1.996 [*]	1.911	1.993		28.8	5
35	5	36.1	2.035	2.054	1.999	2.073	2.294	14.3	5
36	5	37.3	2.289	2.627 [*]	2.071	1.923	2.018	16.3	5
37	5	38.1	2.142	2.088	2.003	2.054	2.227 ^{**}	21.0	5
38	5	39.4	2.041	2.142	1.984	2.058	2.287	55.6	4

39	5	41.6	5.610	1.995	1.969	2.038	2.033	36.9	6
40	5	42.4	2.289	2.267 *	2.020	1.940	2.035	16.3	3
41	5	50.6	2.165	2.306	2.000	2.023	2.025	20.6	3
42	5	51.2	3.889	1.993	1.983	2.053	2.222	58.3	3

* The oxygen atom coordinated to Cu²⁺ does not belong to the COO⁻ group of the Glu side chain residue;

** Histidine in apical position;

Table 2. TDDFT BP86/TZVP electronic spectrum of **1**. For each transition the excitation energy (nm), the oscillation strength (f) and the main ($a_{i \rightarrow i}^{23}$) mono-electronic excitations with the corresponding percentage composition, are reported.

Transition	nm	f	$a \rightarrow i$
1	1005	$7 \cdot 10^{-4}$	202 \rightarrow 203
2	947	$1 \cdot 10^{-3}$	201 \rightarrow 203
3	808	$4 \cdot 10^{-3}$	200 \rightarrow 203
4	757	$2 \cdot 10^{-3}$	199 \rightarrow 203
5	702	$5 \cdot 10^{-3}$	198 \rightarrow 203 (55%) 197 \rightarrow 203
6	689	$1 \cdot 10^{-2}$	(45%) 196 \rightarrow 203
7	685	$2 \cdot 10^{-3}$	196 \rightarrow 203
8	637	$2 \cdot 10^{-4}$	195 \rightarrow 203
9	623	$1 \cdot 10^{-3}$	194 \rightarrow 203
10	597	$2 \cdot 10^{-4}$	193 \rightarrow 203
11	588	$5 \cdot 10^{-3}$	192 \rightarrow 203
12	575	$4 \cdot 10^{-3}$	191 \rightarrow 203
13	552	$6 \cdot 10^{-3}$	190 \rightarrow 203
14	507	$2 \cdot 10^{-3}$	189 \rightarrow 203
15	497	$5 \cdot 10^{-4}$	188 \rightarrow 203
16	485	$5 \cdot 10^{-4}$	187 \rightarrow 203
17	462	$9 \cdot 10^{-4}$	185 \rightarrow 203
18	457	$4 \cdot 10^{-3}$	186 \rightarrow 203 (87%)

Table 3. Selected geometrical parameters (in Å and degree) and energy differences relative to the lowest energy form (in kcal·mol⁻¹) of the [Cu(Ac-PHRQN-NH₂)(H₂O)]⁺¹ complexes. In the table are reported the five Cu-L distances, the dihedral angle that defines the base of the square pyramid of the N₃O₂ coordination sphere and the number of hydrogen bonds (H···O distance less than 2.0Å). CN stands for coordination number.

Isomer	CN	ΔE	Distances					∠ dihedra I	H- bonds
			Cu- O _{H2O}	Cu-O _{CO}	Cu-N _I	Cu-N ₂	Cu-N _{His}		
Q1	5	0.0	2.198		2.278	1.969	2.044	2.024	25.2
Q2	5	1.3	2.301		2.341	2.013	1.973	2.001	20.4
Q3	4	2.2	2.124			1.992	2.001	2.027	16.6
Q4	5	7.1	2.257		2.348	2.049	1.954	1.997	29.5
Q5	4	10.3	2.174			1.997	1.947	1.992	40.8
Q6	5	11.0	2.242		2.252	2.022	1.994	2.012	39.4
Q7	4	13.2	3.495		2.201	1.928	1.992	1.976	37.5
Q8	5	15.2	2.484	2.317		2.026	1.970	2.035	21.4
Q9	5	15.2	2.537	2.302		1.966	2.023	2.031	20.7
Q10	4	16.8	4.652		2.134	1.998	1.931	1.968	29.6
Q11	5	17.9	2.738		2.203	2.060	1.932	1.978	16.3
Q12	3	19.2	4.123			1.972	1.915	1.939	
Q13	4	20.1	4.199	2.273		1.955	2.000	1.998	9.0
Q14	4	21.9	3.668	2.305		1.953	2.009	1.981	48.0
Q15	4	37.4	3.544		2.108	1.957	2.007	2.108	6.1
Q16	4	37.4	3.521		2.107	2.006	1.958	1.983	5.7
Q17	4	39.3	3.181		2.176	2.023	1.948	1.973	19.3

Q18	4	39.4	3.310	2.179	2.023	1.946	1.977	19.1
Q19	4	39.7	2.168		1.962	1.991	1.983	32.7
Q20	4	40.4	3.539	2.172	2.008	1.952	1.989	14.7

Table 4. EPR hyperfine coupling constants (MHz) for the Cu atom in relevant isomers of the [Cu(Ac-PHREN-NH₂)(H₂O)] complex, computed using the B3LYP functional on geometries optimized at the RI-BP86/TZVP level of theory.

Complex	Cu-O _{H2O}	A _{FC} ^a	A _{PC} ^a	A _{iso}	A _{dip} ^{a,b}	A _{dip2} ^{a,b}	total dipolar ^b	A ^c	A _{xx} ^c	A _{yy} ^c	\widehat{A}_{dip} ^d
1	2.645	-308	139	-170	-488; 243; 245 114; -58; -57	-373; 186; 189 -543	16	19	3		
1w	2.445	-303	144	-159	-489; 244; 245 118; -61; -57	-371; 183; 188 -530	25	29	4		
2	3.777	-315	132	-183	-478; 236; 241 113; -54; -60	-364; 183; 181 -547	0	-1	-1		
3	3.863	-319	128	-191	-478; 238; 240 114; -55; -58	-364; 183; 182 -555	-8	-9	1		
4	3.796	-310	127	-183	-478; 240; 238 112; -58; -55	-366; 182; 183 -549	-1	0	1		
5	3.565	-251	140	-111	-456; 203; 252 115; -50; -65	-340; 153; 187 -451	42	76	34		
6	3.374	-291	135	-155	-461; 230; 232 112; -61; -52	-349; 169; 180 -504	13	25	12		
7	3.359	-320	131	-189	-468; 249; 219 113; -61; -51	-355; 188; 168 -544	-1	-21	20		
8	2.488	-286	145	-141	-495; 239; 256 119; -58; -62	-376; 182; 194 -517	41	53	12		
9	3.684	-321	130	-190	-479; 243; 235 114; -58; -55	-366; 185; 180 -556	-5	-10	-5		
10	3.655	-244	135	-109	-432; 171; 262 109; -45; -47	-323; 126; 197 -432	16	88	72		

11	2.492	-276	143	-133	-491; 241; 250 117; -57; -59	-374; 184; 191 -507	51	58	7
16	3.755	-313	134	-179	-476; 238; 239 115; -56; -60	-361; 182; 179 -540	3	0	-3
20	2.407	-230	152	-78	-468; 107; 361 111; -14; -100	-357; 93; 261 -435	17	184	167

Expt.^c

462

^a A_{FC} and A_{dip} are non relativistic Fermi contact and traceless dipolar coupling tensor, respectively. A_{PC} and A_{dip2} are the corresponding relativistic corrections.

^b The components are reported in the order A_{33} , A_{11} , A_{22} .

^c $\hat{A}_{||} = A_{iso} + A_{33}$; $\hat{A}_{xx} = A_{iso} + A_{11}$; $\hat{A}_{yy} = A_{iso} + A_{22}$.

^d \hat{A}_{dip} is computed as $\hat{A}_{dip} = A_{22} - A_{11}$ and may be considered as the deviation from the axial symmetry of the tensor.

^e Experimental values taken from Ref. [16].

Table 5. g tensors of relevant isomers of the [Cu(Ac-PHREN-NH₂)(H₂O)] complex, computed using the B3LYP functional on the geometries optimized at the RI-BP86/TZVP level of theory

	g_{iso}	g_{xx}	g_{yy}	g_{\parallel}
1	2.098	2.055	2.058	2.182
1w	2.103	2.058	2.061	2.190
2	2.092	2.048	2.053	2.175
3	2.090	2.047	2.051	2.171
4	2.088	2.046	2.050	2.170
5	2.097	2.050	2.058	2.184
6	2.095	2.049	2.055	2.180
7	2.091	2.047	2.055	2.172
8	2.106	2.059	2.060	2.192
9	2.091	2.047	2.053	2.172
10	2.093	2.044	2.058	2.177
11	2.102	2.058	2.060	2.189
16	2.093	2.049	2.054	2.177
20	2.109	2.036	2.102	2.190
Expt. ^a				2.238

^a Experimental values taken from Ref. [16].

Table 6. EPR hyperfine coupling constants (MHz) for Cu atoms in relevant isomers of the $[\text{Cu}(\text{Ac-PhRQN-NH}_2)(\text{H}_2\text{O})]^{+1}$ complex, computed using the B3LYP functional on geometries optimized at the BP86/def-TZVP level of theory.

Complex	Cu-O _{H2O}	A _{FC} ^a	A _{PC} ^a	A _{iso}	A _{dip} ^{a,b}	A _{dip2} ^{a,b}	total dipolar ^b	A ^c	A _{xx} ^c	A _{yy} ^c	\widehat{A}_{dip} ^d
Q1		-206	144	-62	-452; 81; 371	111; 11; -100	-341; 70; 271	-403	8	109	101
Q2		-234	145	-89	-465; 123; 342	107; -21; -86	-358; 102; 256	-447	13	167	154
Q3		-319	133	-186	-471; 235; 236	109; -53; -56	-362; 182; 180	-548	-4	-7	-3
Q4		-120	149	29	406; 30; -436	-113; 9; 104	293; 39; -332	322	68	-304	-372
Q5	2.174	-255	126	-129	-450; 196; 254	98; -42; -57	-352; 154; 198	-481	25	69	44
Q6	2.242	-253	128	-125	-441; 80; 362	101; -11; -90	-317; 69; 272	-465	-56	147	203
Q7	3.495	-214	124	-90	-443; 178; 265	97; -37; -60	-346; 142; 205	-436	52	115	63
Q8		-234	144	-90	-461; 215; 246	111; -50; -60	-350; 165; 186	-441	76	96	-20
Q11		-165	131	-34	-456; 167; 289	110; -36; -72	-346; 131; 217	-380	97	183	96
Q16		-251	124	-127	-462; 214; 248	109; -50; -59	-353; 165; 188	-480	28	68	40
Expt. ^c									489		

- ^a A_{FC} and A_{dip} are non relativistic Fermi contact and traceless dipolar coupling tensor, respectively. A_{PC} and A_{dip2} are the corresponding relativistic corrections.
- ^b The components are reported in the order A_{33} , A_{11} , A_{22} .
- ^c $A_{\parallel} = A_{\text{iso}} + A_{33}$; $A_{\text{xx}} = A_{\text{iso}} + A_{11}$; $A_{\text{yy}} = A_{\text{iso}} + A_{22}$.
- ^d \hat{A}_{dip} is computed as $\hat{A}_{\text{dip}} = A_{22} - A_{11}$ and may be considered as the deviation from the axial symmetry of the tensor.
- ^e Experimental values taken from Ref. [16].

Table 7. g tensors of relevant isomers of the $[\text{Cu}(\text{Ac-PHRQN-NH}_2)(\text{H}_2\text{O})]^{+1}$ complex, computed using the B3LYP functional on geometries optimized at the RI-BP86/TZVP level of theory.

	g_{iso}	g_{xx}	g_{yy}	g_{\parallel}
Q1	2.106	2.023	2.100	2.184
Q2	2.102	2.036	2.088	2.181
Q3	2.091	2.048	2.055	2.171
Q4	2.102	2.030	2.094	2.182
Q5	2.085	2.041	2.056	2.159
Q6	2.090	2.026	2.083	2.161
Q7	2.084	2.037	2.058	2.157
Q8	2.098	2.051	2.063	2.179
Q11	2.091	2.039	2.065	2.169
Q16	2.084	2.043	2.050	2.160
Expt. ^a				2.240

^a Experimental values taken from Ref. [16].

Figure 1. Lowest energy structure of the $[\text{Cu}(\text{Ac-PHREN-NH}_2)(\text{H}_2\text{O})]$ complex characterized by the 2-3 binding mode (1). Distances in Å.

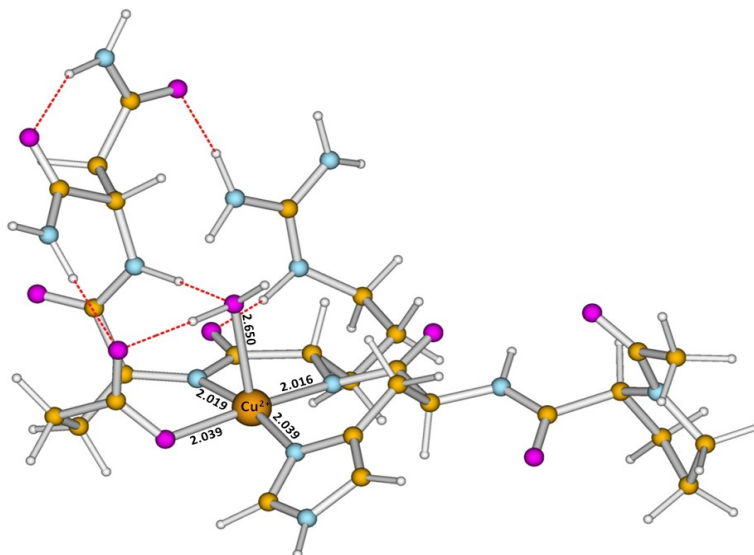


Figure 2. TDDFT computed UV-Vis spectra for Ac-PHREN-NH₂ and the lowest energy structure of the [Cu(Ac-PHREN-NH₂)(H₂O)] complex (**1**)

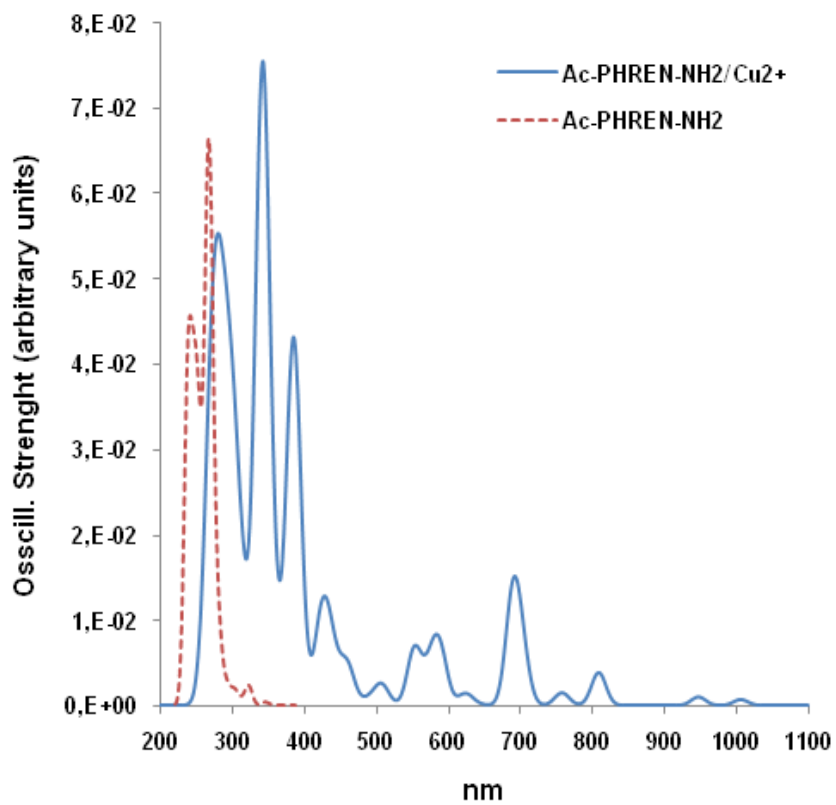


Figure 3. Lowest energy structure of the $[\text{Cu}(\text{Ac-PHREN-NH}_2)(\text{H}_2\text{O})]$ complex characterized by the 1-2 binding mode (**21**). Distances in Å.

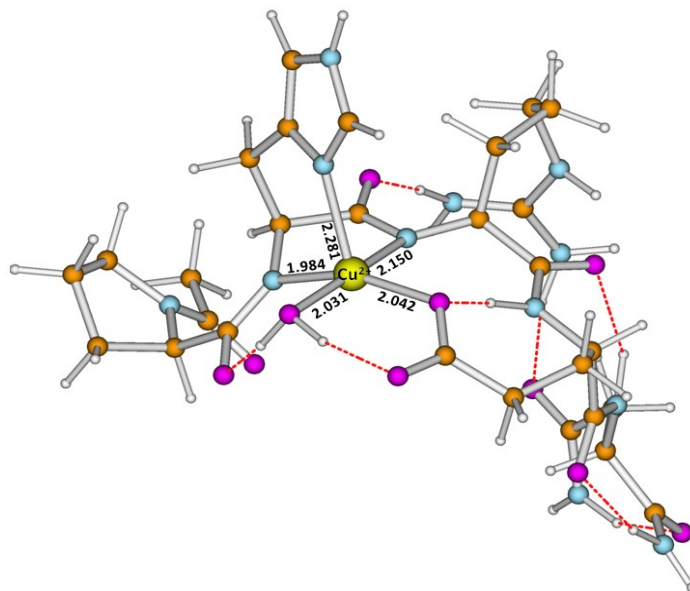
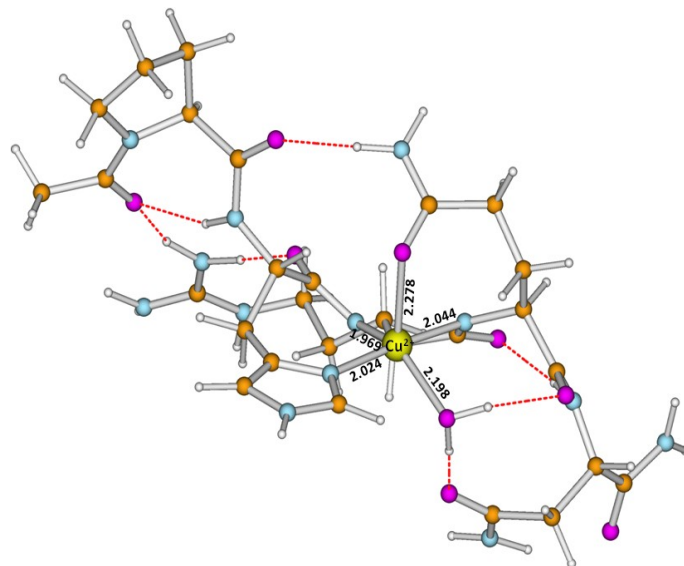


Figure 4. Lowest energy structure of the $[\text{Cu}(\text{Ac-PHRQN-NH}_2)(\text{H}_2\text{O})]^{+1}$ complex (**1Q**). Distances in Å.



DFT Characterization of Key Intermediates in Thiols Oxidation Catalyzed by Amavadin

Luca Bertini, Valentina Barbieri, Piercarlo Fantucci, Luca De Gioia, Giuseppe Zampella* Department of Biotechnology and Biosciences, Università degli Studi di Milano-Bicocca, Piazza della Scienza, 2; 20126 Milan - Italy

†To whom correspondence should be addressed: giuseppe.zampella@unimib.it

Abstract

Amavadin is an unusual octa-coordinated V^{IV} complex isolated from *Amanita muscaria* mushrooms. The outer-sphere catalytic properties of such complex toward several oxidation reactions are well known. Nevertheless, a remarkable example exists, in which the V^V (d⁰) oxidized form of amavadin is able to electro-catalyze the oxidation of some thiols to the corresponding disulfides through an inner-sphere mechanism (Guedes da Silva et al. *J. Am. Chem. Soc.* **1996**, *118*, 7568-7573.) The reaction mechanism implies the formation of an amavadin-substrate intermediate, whose half-life is about 0.3 s. By means of Density functional Theory (DFT) computations and Quantum theory of Atoms in Molecules (QTAIM) analysis of the electron density, we have first characterized the stereoelectronic features of the V^{IV} (inacti-

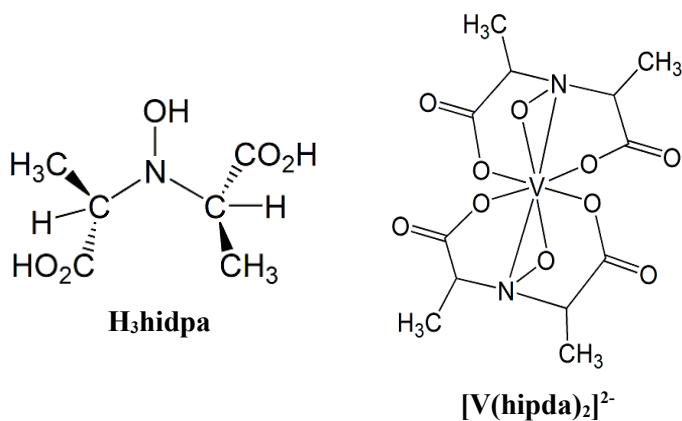
ve) and V^V (active) states of amavadin. Then, the formation of the V^V complex with methyl mercaptoacetate (MMA), which has been chosen as prototypical substrate, has been characterized both thermodynamically and kinetically. DFT results reveal that protonation of V^V amavadin at a carboxylate oxygen not directly involved in the V coordination, favors MMA binding into the first coordination sphere of vanadium, by substitution of the amavadin carboxylate oxygen with that of the substrate and formation of a S-H \cdots O hydrogen bond interaction. The latter interaction can promote SH deprotonation and binding of the thiolate group to vanadium. The kinetic and thermodynamic feasibility of the V^V -MMA intermediates formation is in agreement, along with electrochemical experimental data, also with the biological role exerted by amavadin. Finally, the presence of an ester functional group as an essential requisite for MMA oxidation has been rationalized.

1. Introduction

For its ubiquitous distribution in the entire biosphere and its highly amphoteric oxidative character, vanadium has been found to have so much relevance in biological systems that its role as an essential metal in the first steps of evolution has recently been hypothesized.ⁱ Representing the 0.014% of the lithosphere and the second-most abundant transition metal in sea-water, with a concentration of 30 nM,ⁱⁱ vanadium is widely available to living organisms. Moreover, the ease of shifting between three oxidation states (+3, +4, +5) of biological relevance further emphasizes its

capability of serving as a redox shuttle.

Some species of the genus *Amanita*, e.g. *A.muscaria*, *A.muscaria regalis*, *A.velatipes*, show concentrations of vanadium of up to 400 mg kg⁻¹ in the form of a natural vanadium(IV) compound, called amavadin.^{iii,iv} It is well known that the specific chemical environment of the vanadium centers in natural vanadium complexes and proteins plays a fundamental role in tuning the metal biological activity. In this scenario the very peculiar structure of amavadin has raised growing interest from both a chemical and biological point of view. Amavadin is a 2:1 octa-coordinated complex between V^{IV} and S,S-2-2'-hydroxyimino-dipropionic acid HON{CH(CH₃)COOH} (H₃hidpa) in its deprotonated form, and metal coordination involves the carboxylic and hydroxyamino groups of the ligand:



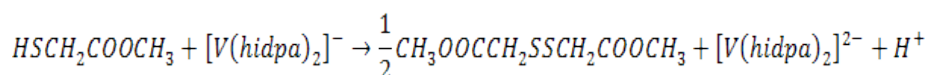
Remarkably, before isolating it in amavadin crystal structure, H₃hidpa ligand was completely unknown. The crystallographic characterization of amavadin^v ([V(hidpa)₂]²⁻ hereafter) revealed that no vanadyl group

[VO]²⁺ is present, as opposed to precedent hypothesis.^{vi} The complex displays an intriguing *non-oxo* octacoordinated V^{IV} ion with a distorted dodecahedral {VO₆N₂} coordination, with two η² N-O of the oxyimino group and four unidentate carboxylate groups. The relative orientation of the two N-O leads to the Δ e Λ chirality of the metal center. In fact, spectroscopic studies^{Error: sorgente del riferimento non trovata} have revealed that isolated amavadin consists of an equimolar racemic mixture of two stereoisomers, Δ e Λ.

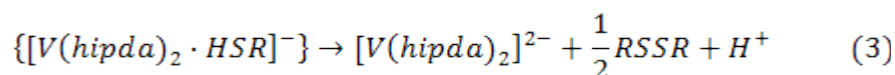
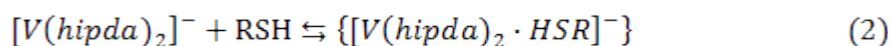
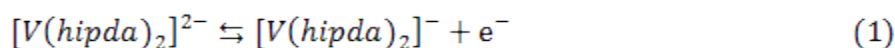
For its reversible one-electron redox properties, amavadin can act as a catalyst for thiols oxidation,^{vii} peroxidative halogenation, hydroxylation and oxygenation of alkyl and aromatic substrates^{viii}. Recently Guedes da Silva et al. reported that amavadin can also convert methane into acetic acid^{ix} and ethane into propionic and acetic acid.^x In light of these findings, it is plausible to define amavadin as a sort of proto-enzyme even though its real function has not yet been identified.^{Error: sorgente del riferimento non trovata}

Hubretgste et al.^{xi} demonstrated that the carboxylate groups are essential for the stability of amavadin and consequently an outer rather than inner sphere catalytic mechanism for a series of electrocatalytic transfer reactions was initially hypothesized. Notwithstanding, prompted by a work by Tackerey et al.^{xii}, Guedes da Silva et al.^{xiii} gave experimental evidences of an inner sphere mechanism in the oxidation of thiols (RSH) to the corresponding disulfides (RSSR) by amavadin. The latter reaction, which could be involved in the protective/defensive system of the mushroom since it could regenerate broken disulfides bridges in protein fibers of damaged tissues,^{Error:}

sorgente del riferimento non trovata involves the one-electron oxidized V^V form of amavadin with the formation of the corresponding disulfide.



In particular, the following reaction steps were proposed on the basis of cyclic voltammetry measurements:

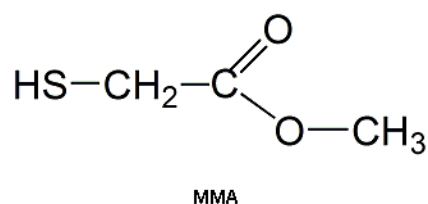


This redox catalysis process occurs through a Michaelis-Menten kinetic mechanism, as revealed by saturation of the catalytic efficiency upon increasing substrate concentration. The active catalytic species is the $d^0 V^V$ amavadin and the transient formation of an intermediate species $[V(hidpa)_2 \cdot HSR]^{-}$ generated by coordination of the oxidized vanadium complex to the thiolic substrate has been observed. In particular, Guedes da Silva et al. found a rate constant for the forward reaction (2) of $1.2 \cdot 10^3 \text{ M}^{-1} \text{ s}^{-1}$, and estimated a half life of 0.3 s for the $[V(hidpa)_2 \cdot HSR]^{-}$ intermediate, when mercaptoacetic or mercaptopropionic acid were used as substrates.

Quantum chemical approaches are useful tools in the characterization

of the stereoelectronic properties of coordination compounds. In fact, the electronic structure of amavadin in its d^1 reduced and oxidized d^0 forms was investigated at DFT level in a few previous theoretical works. Armstrong and co-workers^{xiv} presented a $X\alpha$ investigation in order to elucidate character of the frontier MOs, showing the similarity of the $d_{x^2-y^2}^{(1)} d_{yz} d_{xz} d_{yz} d_{z^2}$ d-orbital ordering of $[V(\text{hidpa})_2]^{2-}$ with that of oxovanadium(IV) VO^{2+} complexes, as $[\text{VOCl}_4]^{2-}$. In light of these results, the electronic spectrum of the reduced amavadin form was computed and compared with the experimental one. Remenyi and co-workers^{xv} and Ooms and co-workers^{xvi} dealt with the computation of EPR and ^{51}V NMR parameters of the reduced form of amavadin, while in a more recent paper Geethalakshmi^{xvii} and co-workers investigated the ^{51}V NMR parameters of the oxidized form.

In the present work we have used Density Functional Theory (DFT) to study the reactivity of amavadin with methyl mercaptoacetate (MMA),



which is known to undergo oxidation to the corresponding disulfide, according to the mechanism proposed by Guedes da Silva et al.^{Errore: sorgente del riferimento non trovata} Then, the electronic structure and the bonding properties of key specie involved in the catalytic process

were investigated by means of the Quantum Theory of Atom in Molecules approach (QTAIM), which was never used before to study the stereoelectronic features of amavadin.

The present study also fits into the line of investigations about the controversial vanadate reduction by thiols. A recent work by Crans et al.^{xviii} shows that both V^{IV} and V^V can form complexes with thiols and that the redox chemistry that can occur is governed by different parameters such as pH, solute- and vanadate-concentrations, as well as the presence of other ligands. The present investigation has thus taken into account the possibility of forming a methyl mercaptoacetate-amavadin adduct through V^V-S bond making. The role of an ancillary (respect to thiol) functional group, such an ester, has been elucidated.

2. Computational details

Computations were performed using the pure Generalized Gradient Approximation (GGA) B-P86 DFT functional^{xix} and the Resolution of Identity^{xx} technique, as implemented in the TURBOMOLE^{xxi} suite of programs. Basis sets of triple-zeta plus polarization split valence quality (TZVP hereafter)^{xxii} were adopted for all atoms in the complexes. The DFT grid-size was set to standard m3 value. The employed theory level has turned out^{xxiii,xxiv} to be well suited for DFT investigations regarding reactivity properties of vanadium(V) complexes, which are closely related to amavadin, such as the cofactor of vanadium haloperoxidase (VHPO). Solvent effects were taken into account using the COSMO (Conductor-like Screening

Model; $\epsilon=80.0$) approach.^{xxv,xxvi,xxvii} The pseudo-Newton-Raphson procedure has been adopted for the search of transition states (TS). Schematically, it entails the structure optimization (SO) of a guess of the true TS. During this SO, those degree of freedom which are supposed to vary most along the reaction coordinate (RC) are kept frozen; this step allows all the structure but the reactive core to reach a potential energy minimum. After calculating the Hessian matrix associated to the optimized guess TS, (the presence of a negative eigenvalue corresponding to the motion along the RC ensures about the nature of the found saddle point) an eigenvector-following procedure is adopted for locating the maximum energy point along the RC, i.e., the TS.

QTAIM analysis of the electron density was implemented using AIMPAC^{xxviii} and SF_ESI^{xxix} codes. Here, it is worth briefly introducing the definition and meaning of the delocalization index $\delta(A,B)$. This index is an integral property that indicates the number of electron pairs delocalized between two atoms, and can be considered a covalent bond order. In the case of homopolar pure covalent bonds as in H₂ or in the C-C bond in ethane, the delocalization index value is 1, while in NaF, where the ionic contribution to the bond is significant $\delta(\text{Na},\text{F})$ is 0.27.^{xxx,xxxi} Delocalization indexes have been recently employed to study organometallic systems^{xxxii,xxxiii} and turned out to be valuable parameters for the description of the metal-metal and metal-ligand bonds, highlighting their non-localized nature.

Calculations have been carried out only on the Δ stereoisomer of amavadin, so in the following the symbol Δ will be omitted for

nomenclature simplification.

3. Results and discussions

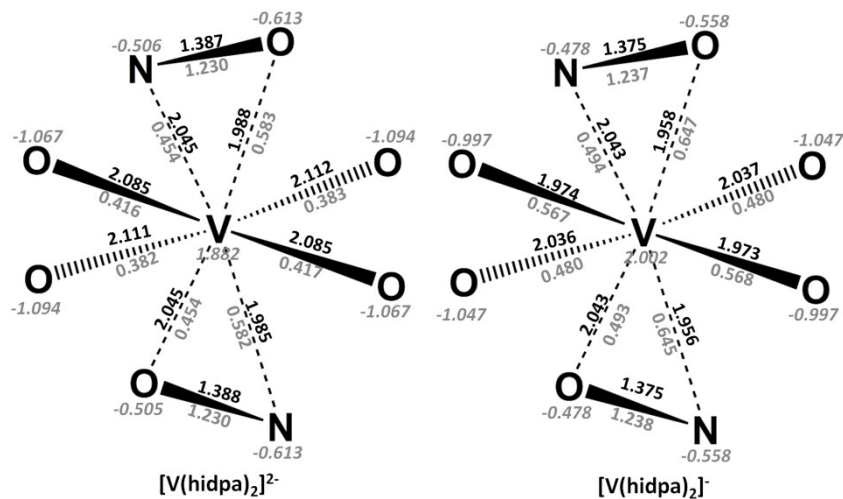
To shed light on the structural features of key intermediate species formed in the inner sphere mechanism of thiol oxidation experimentally studied by Guedes da Silva et al.,¹ the coordination mode of a prototypical substrate (MMA) to the vanadium atom of amavadin has been studied using DFT. In fact, experimental evidences showed that the presence in the substrate molecule of a carboxylic derivative two or three bonds away from the sulfhydryl group is necessary for catalysis to take place. Moreover, both H-S-R-COOR' thiols with R' corresponding to carboxylate or ester groups undergo oxidation, indicating that the hydroxylic group of the carboxyl moiety of the substrate does not play a fundamental role in the reaction. According to this observation, MMA rather than mercaptoacetic acid has been chosen as the substrate in our calculations.

To shed light on the reaction mechanism, the possibility of a protonation event on amavadin has been also taken into account, since catalysis takes place in a protic KCl/H₂O electrolyte solution.¹ Therefore, both protonated and non-protonated forms of oxidized amavadin have been considered.

3.1 Geometry and Electronic structure of reduced and oxidized forms of amavadin

The V^{IV} $[V(\text{hidpa})_2]^{2-}$ structure has been optimized starting from its X-ray crystallographic structure, while the V^V $[V(\text{hidpa})_2]^-$ structure has been optimized starting from the optimized $[V(\text{hidpa})_2]^{2-}$ atomic coordinates. The discussion of the electronic spectra of the two species is presented in the supplementary materials.

The vanadium coordination in $[V(\text{hidpa})_2]^{2-}$ and $[V(\text{hidpa})_2]^-$, as well as DFT optimized bond distances, $\delta(\text{A,B})$ delocalization indexes, and QTAIM atomic net charges are shown in Figure 1. On average, the $[V(\text{hidpa})_2]^{2-}$ optimized V-O and V-N bond distances are overestimated by 0.030\AA , with respect to X-ray diffraction values.^{xxxiv} The same comparison made for optimized $[V(\text{hidpa})_2]^-$ bond distances revealed a similar average error (0.034\AA).



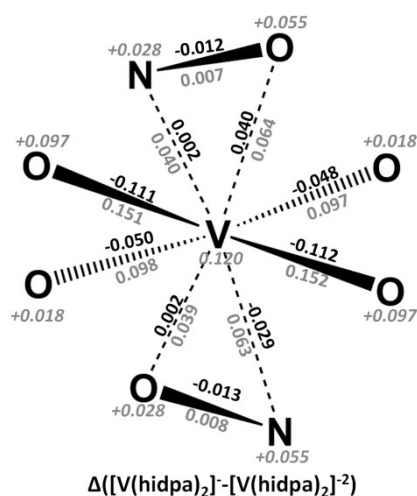


Figure 1. From left to right: $[\text{V}(\text{hidpa})_2]^{2-}$ and $[\text{V}(\text{hidpa})_2]^{-}$ $\{\text{VO}_6\text{N}_2\}$ first vanadium coordination sphere optimized bond distances (in black, in Å), delocalization indexes (grey) and atomic net charges (grey, italics). In the last picture are reported the $\text{V}^{\text{V}}-\text{V}^{\text{IV}}$ specie differences among the properties computed. $[\text{V}(\text{hidpa})_2]^{2-}$ X-ray bond distances: V-NO 1.999(8)Å, 1.982(8)Å; V-ON 1.940(7)Å, 1.956(7)Å; V-O 2.070(8)Å, 2.028(7)Å; 2.028(9)Å, 2.042(8)Å; 2.070(8)Å (Ca^{2+} counterions, see ref. XX); $[\text{V}(\text{hidpa})_2]^{-}$ X-ray bond distances: V-NO 1.999Å, 2.018Å; V-ON 1.940Å, 1.923Å; V-O 1.993Å, 1.977Å; 1.972Å, 1.959Å; (PPh_4^+ counterions see ref XX)

The oxidation of $[\text{V}(\text{hidpa})_2]^{2-}$ to $[\text{V}(\text{hidpa})_2]^{-}$ releases $13.7 \text{ kcal}\cdot\text{mol}^{-1}$ but does not modify significantly the structure of the complex, in agreement with experimental data. However, a contraction of the V-O(1) and V-O(5) bonds (5.3-5.4%) and of the V-O(6) and V-O(10) bonds (3.7%) occur, whereas the V-O(3) and V-O(8) bonds show a modest contraction (1.5%) and the V-N bond a minimal shrinkage (0.1%).

The $\delta(\text{V-X})$ delocalization indexes are useful parameters for a quantitative estimation of the strength of the covalent component of a given

chemical bond. In $[\text{V}(\text{hidpa})_2]^{2-}$ V-X bonds involve 3.7 electron pairs. The V-ON bond results the strongest in the $\{\text{VO}_6\text{N}_2\}$ group, and summing $\delta(\text{V-X})$ for V-ON and V-NO bonds, we conclude that V- η^2 -NO bonding involves 1.035 electron pairs, which corresponds to a single covalent bond. The weakest V-O bonds (V-O(10) and V-O(6); Figure 1) involve 0.38 pairs. As expected, upon oxidation, V-X bonds become shorter and stronger with respect to the reduced form, according to the decreased bond distances and the increased $\delta(\text{V-X})$ values.

The effect of the one-electron oxidation on the electronic structure of $[\text{V}(\text{hidpa})_2]^{2-}$ depends on which property is discussed. The increase of the atomic net charges is essentially distributed over the whole systems. The total increase of the atomic net charges of the V atom and its $\{\text{VO}_6\text{N}_2\}$ first coordination sphere is 0.52 (see Figure 2, right), where V positive charge increases by 0.12 e. However, the spin population of $[\text{V}(\text{hidpa})_2]^{2-}$, which is completely localized on the V atom ($\alpha-\beta=1.011$ e), disappears in the closed-shell $[\text{V}(\text{hidpa})_2]^-$. This last point is also evidenced in the analysis of the population of the $[\text{V}(\text{hidpa})_2]^{2-}$ singly occupied HOMO, which is dominated by the V *d*-orbitals contributions.

3.2 DFT optimization of the protonated form of $[\text{V}(\text{hidpa})_2]$

The kinetics data Errore: sorgente del riferimento non trovata reported by Guedes da Silva et al. show that no pH effects were detected on the electrocatalytic activity of amavadin. It was proposed that this

behavior could depend on the high stability constants of the amavadin complex ($>10^{21}$), implying that the catalyst remains undissociated even at very low pH values. This observation is apparently in contrast with the proposed inner sphere mechanism, since if the chelate-V bonds are too strong to be cleaved, formation of a vacant site for MMA coordination would be prevented. In addition, it must be also noted that substitution of one of the carboxylate groups of the hidpa ligand with the ester group of MMA would not be energetically favorable.

In this scenario, it is conceivable that protonation of one vanadium ligand in amavadin can take place and weaken one of the V-X bonds, thus favoring the coordination of the substrate to the active redox form of the catalyst. Prompted by this observation, the $[\text{V}^{\text{V}}(\text{hidpa})_2]^-$ structure has been subjected to a systematic sampling of all the possible protonation sites, with the aim of locating the most basic sites (according to Brønsted-Lowry definition) of the V complex. Each oxygen atom not included in the first V coordination sphere has been initially protonated considering the two different possible orientations of the H atom with respect to V (in structures A and B the H atom points toward or in opposition to the V atom, respectively). Similarly, both “up/down” orientations (relative to the equatorial plane of the complex) have been considered for protonation at those O atoms which are ligated to V. The protonated forms of amavadin will be referred to as $[\text{V}(\text{hidpa})_2\text{H}]$.

Relative stability data for all the protonated forms investigated, as well as the V-O and V-N bond distances for the four more stable

isomers, are reported in Table 1.

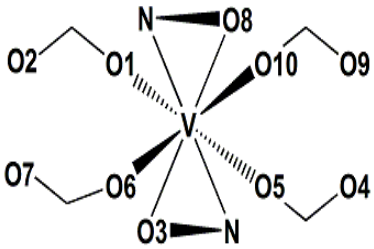
	Protonation at	ΔE	$d(\text{V-OH})$		
	O1_A	5.123	2.213		
	O1_B	5.121	2.216		
	O2	0.000	2.197		
	O3_A	7.975	2.977		
	O3_B	7.987	2.976		
	O4	0.011	2.199		
	O5_A	5.117	2.213		
	O5_B	5.115	2.213		
	O6_A	7.377	2.399		
	O6_B	7.358	2.397		
	O7	0.268	2.034		
	O8_A	8.000	2.971		
	O8_B	7.984	2.971		
	O9	0.318	1.955		
	O10_A	7.364	2.392		
	O10_B	11.424	3.213		
	[V(hidpa)₂H]				
	(O)2 (O)4 (O)7 (O)9 (O)10				
Distances					
V-O(3)	1.940	2.199	1.975	2.188	1.915
V-O(6)	1.950	1.935	1.960	1.930	1.959
V-O(8)	1.956	1.998	2.034	2.000	1.915
V-O(10)	1.932	1.950	1.956	1.955	3.213
V-O(5)	1.998	1.957	2.036	1.962	1.920
V-O(1)	2.197	1.939	1.972	1.996	1.984
V-N(1)	2.051	2.033	2.043	2.035	2.100
V-N(2)	2.035	2.050	2.043	2.051	2.124
ΔE	0.0	0.01	0.27	0.3	11.42

Table 1. Properties of the protonated $[\text{V}(\text{hidpa})_2\text{H}]$ isomers. ΔE in $\text{kcal}\cdot\text{mol}^{-1}$ with respect to the most stable isomer; $d(\text{V-OH})$ is the optimized bond distance between V and the protonated O atom; in the second part of the table are reported the V-O and V-N optimized bond distances of the four most stable isomers. The label A and B refer to the initial orientation of the proton with respect to the metal center

(see text for details). All distances are in Å.

It turned out that amavadin protonation is always favored at the oxygen atom not directly bound to V. In fact, the four structures protonated respectively at O(2), O(4), O(7) and O(9) are essentially isoenergetic, with energy differences that span in the narrow range of 0-0.3 kcal·mol⁻¹. The other protonated structures differ in energy by at least 5.1 kcal·mol⁻¹ from this group. Note that in most cases, the starting point structure of type A and B converged to the same minimum.

The structure of the two [V(hidpa)₂H] isomers, which differ in energy by 5.1 kcal·mol⁻¹, is shown in Figure 2.

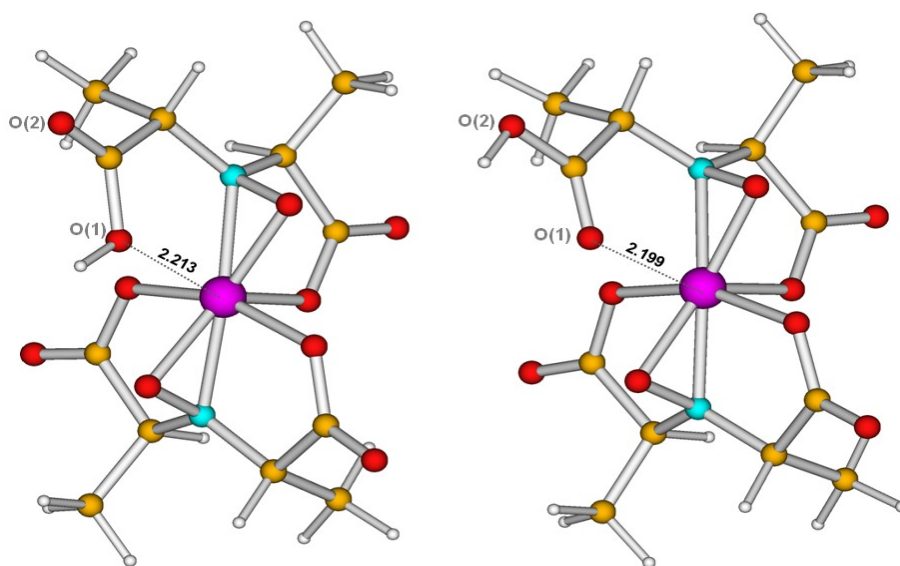


Figure 2. Optimized geometries of V^V [V(hidpa)₂H] isomers in which O(1) (left) and O(2) (right) is protonated. The V-O distances in Å. Vanadium is in yellow, nitrogen in cyan, oxygen in red, carbon in

orange and hydrogen in white.

Clearly the influence on $[\text{V}(\text{hidpa})_2\text{H}]$ geometry of protonation at O(2) is lower than that at O(1). In particular, the $\{\text{VO}_6\text{N}_2\}$ coordination is perturbed by the lengthening of the V-O(1) bond distances (+0.239 Å and, +0.223 Å, with respect to $[\text{V}(\text{hidpa})_2]$, respectively). It is worth noting that except for the atoms directly involved, protonation does not sensibly modify the molecule geometry, since just in one out of the sixteen optimizations that have been carried out a significant change occurs. In this particular case, V-O(10) bond is elongated by 58% and the structure lies 11.4 kcal·mol⁻¹ higher in energy compared to the most stable one.

The QTAIM analysis of the electron density of the structure protonated at O(2) (see Figure 3) evidences that the effect of protonation on the $\{\text{VO}_6\text{N}_2\}$ moiety is essentially localized on the V-O(1) bond. In particular, the V-O(1) bond critical point^{xxxxv} is still found but the $\delta(\text{V-O}(1))$ index is almost halved (-0.31 electron couples, -44%), highlighting the strong weakening but not the dissociation of the V-O(1) bond. All atomic net charges do not change significantly upon protonation, except for that of O(1) that decreases by 0.06 electrons.

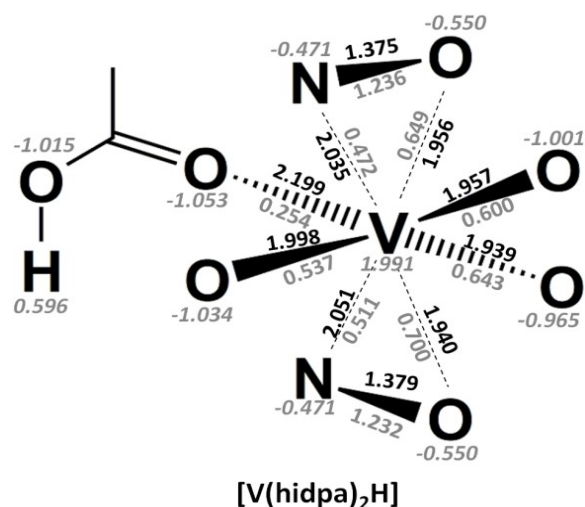
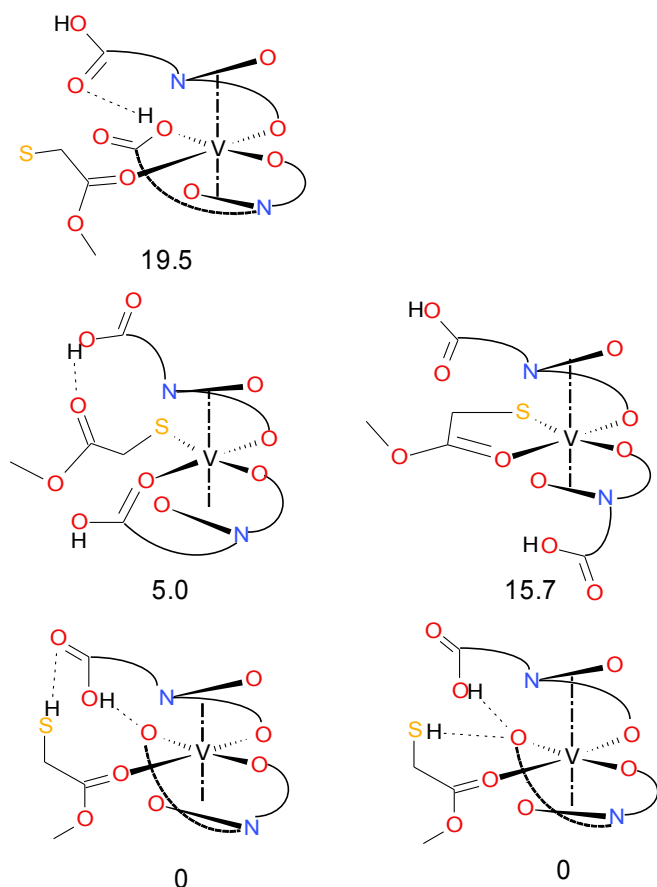


Figure 3. Protonated [V(hidpa)₂H] {VO₆N₂} first vanadium coordination sphere optimized bond distances (in black, in Å), delocalization indexes (grey) and atomic net charges (grey, italics).

3.3 Characterization of reaction intermediates (adducts) formed when MMA reacts with [V(hidpa)₂H]

Every plausible coordination of MMA to [V(hidpa)₂H] protonated at O(2), O(4), O(7), O(9) has been considered. In Scheme 1 are sketched all the intermediate structures identified.



Scheme 1. The complete set of adducts between MMA and $[V(\text{hidpa})_2\text{H}]$ is shown. Relative stabilities are indicated in $\text{kcal}\cdot\text{mol}^{-1}$.

The most energetically favorite adduct shows MMA coordination to V^V via the ester group, by substitution of the protonated $V\text{-O}(2)$ bond, with a formation energy value of $2.7 \text{ kcal}\cdot\text{mol}^{-1}$. The relevant structural features of this adduct and of the TS associated to its formation are collected in Table 2.

	TS of ([V(hidpa) ₂ H]·M MA) formation	[V(hidpa) ₂ H]·M MA	
	V- O(MMA))	2.623	2.150
	V-O(3)	1.958	1.958
	V-O(6)	2.005	2.006
	V-O(8)	1.932	1.932
	V-O(10)	1.963	1.963
	V-O(5)	1.932	1.932
	V- O(2)·H	3.615	3.614
	V-N(1)	2.045	2.045
	V-N(2)	2.129	2.129

Table 2. Structure of the [V(hidpa)₂H]·MMA adduct and transition state (TS) which is associated to its formation by starting from reactants (i.e. an intermolecular van der Waals complex which is held by a hydrogen bond between MMA and protonated amavadin). Vanadium is shown in yellow, nitrogen in cyan, oxygen in red, carbon in orange, hydrogen in white, sulphur in green. The interactions between the MMA and amavadin and the new intramolecular H bond are evidenced in heavy dot line. Distances are in Å.

In particular, i) the carbonyl oxygen of MMA becomes part of the {VO₆N₂} coordination sphere, with a V-O(MMA) bond distance of 2.150 Å; ii) the hydrogen atom of the protonated carboxylate O(2) forms a hydrogen bond with O(6); iii) the hydrogen atom of the thiol group forms a hydrogen bond with the carbonyl oxygen of the protonated carboxylate group (H-O distance = 2.061 Å). Other [V(hidpa)₂H·MMA] isomers, featuring a thiolate group and neutral carboxylic groups of hidpa, are at least 8.2 kcal·mol⁻¹ higher in energy. In order to characterize also kinetically the found adduct, a transition

state (TS) along the reaction coordinate of formation has been searched. The located TS structure reported in Table 2 is approximately quite similar to the $[V(\text{hidpa})_2\text{H}\cdot\text{MMA}]$ found and, therefore, the same sketch has been used to describe both structures (of course evidencing that the coordination of MMA to the catalyst occurs after the formation of the H-bond between the protonated O(1) and O(6), since such interaction is detected already in the TS structure. The activation barrier computed with respect to the $[V(\text{hidpa})_2\text{H}]\cdot\text{MMA}$ van der Waals complex is $20.0 \text{ kcal}\cdot\text{mol}^{-1}$, a value which is indicative of the kinetic accessibility of formation reaction. Globally, both thermodynamics and kinetics of the adduct formation are consistent with a relatively facile reaction and therefore support strongly the possibility of an unprecedented example of inner-sphere electrocatalysis turnover by the amavadin.

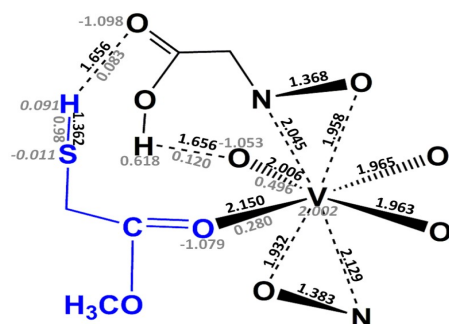
Turning the attention to less stable isomeric forms which has been located concerning the adduct substrate-catalyst, it is worth focusing onto 5.0 and $15.2 \text{ kcal}\cdot\text{mol}^{-1}$ species. Both features S-to-V coordination, the former being stabilized by an intramolecular H bond between ester group on MMA and a non-coordinated carboxylic group whereas the latter being characterized by the formation of an intramolecular 5-membered chelating ring, through the formation of a second V-O(MMA) bond. It is fair to underline that the relative stability which characterizes these two isomers is likely to be overestimated in favor of the mono-coordinated species due to the use of an implicit treatment of solvent effects. Indeed the $15.2 \text{ kcal}\cdot\text{mol}^{-1}$ form, which features two free carboxylic groups, is expected to

establish a number of H bond interactions with water solvent molecules.

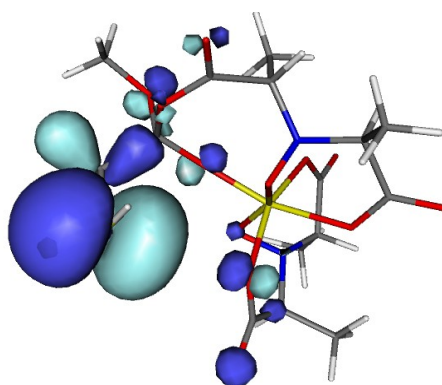
We now discuss the effect of the adduct formation on the electronic structure of the substrate and amavadin by analyzing the QTAIM parameters of the atoms and bonds involved in this step of the reaction (Figure 4). As a general remark, the formation of the intermediate has no dramatic effects on the electronic structure of amavadin, although some small but significant effects on MMA are highlighted.

The formation of the H bond between the hydrogen of the -SH group and the oxygen of the carboxylate ($\text{H}\cdots\text{O}$ distance 1.656 Å) and the corresponding weakening of the S-H bond (the value of $\delta(\text{S,H})$ decreases from 1.042 in free MMA to 0.980 in the intermediate) is able to make the S atom slightly more nucleophilic. Indeed, the initial low but positive net charge of S in free MMA (+0.0167) becomes negative (-0.011), with an increase of the negative charge by 0.027 electrons. This fact is also well evidenced by the shape of the HOMO of the intermediate, that highlights a significant contribution of the sulfur atomic orbitals (see Figure 4, bottom view).

The V-O(MMA) interaction is weak, with a $\delta(\text{V-X})$ delocalization index of only 0.28 electron pairs, which is less than half of the value for other V-O interactions with the carboxylate of the hidpa ligands.



[V(hidpa)₂H·MMA]



HOMO

Figure 4. Top: schematic view of the [V(hidpa)₂H·MMA] intermediate with {VO₆N₂} first vanadium coordination sphere (in black) optimized bond distances (in black, in Å), delocalization indexes (grey) and atomic net charges (grey, italics). MMA is highlighted in blue. Bottom: iso-surface (0.03) plot of the [V(hidpa)₂H·MMA] HOMO.

3.4 Characterization of reaction intermediates formed when MMA reacts with [V(hidpa)₂]

The stereoelectronic characterization of key intermediates formed in the reaction between amavadin [V(hidpa)₂]⁻ and MMA is crucial to cast light on this unprecedented inner-sphere oxidation mechanism, operated by an otherwise outer-sphere performing catalyst.

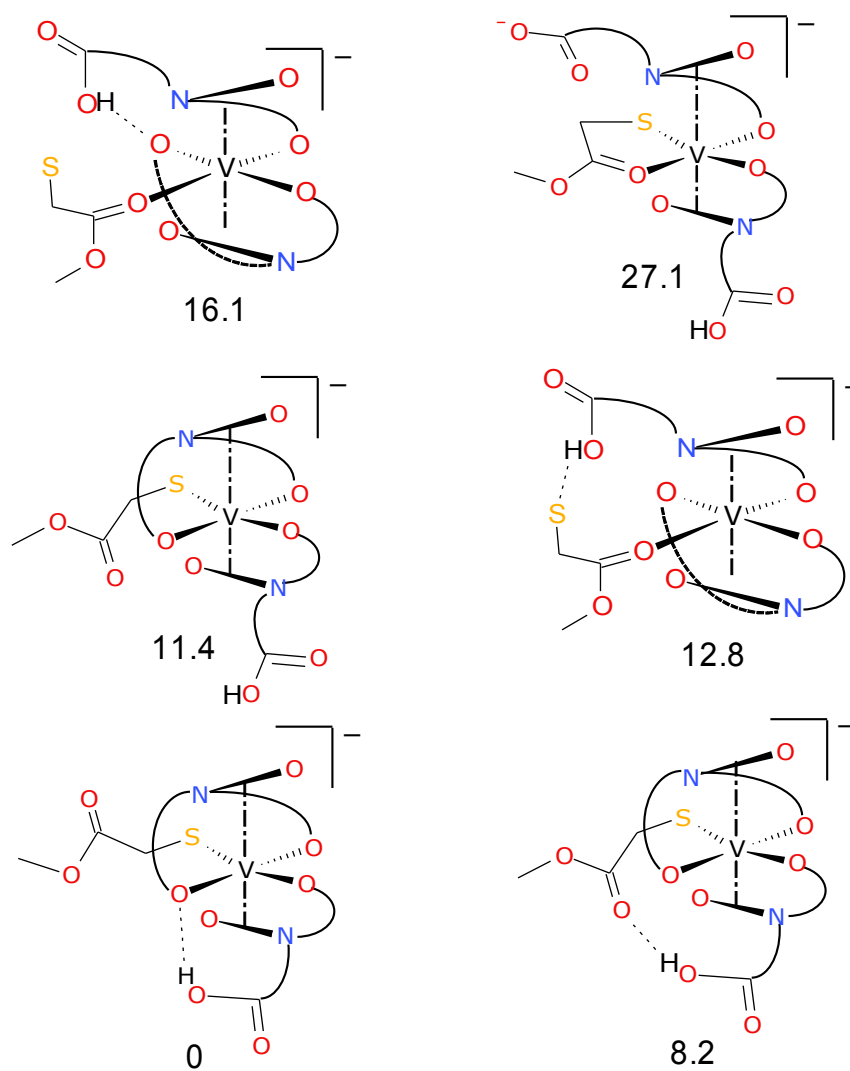
As shown in the previous section, all the plausible coordination of MMA to [V(hidpa)₂H] have been considered, and the different isomers characterized by DFT, as well as their relative energies (Scheme 1).

In light of the results published by Crans et al¹, which demonstrated that mercapto-aminoacid like cystein can bind to V^V in a non-reductive way through the thiol group, initially the possibility has been considered of forming the MMA-amavadin adduct by a V-S bond. However, structures characterized by coordination of the SH group to vanadium do not correspond to energy minima. To explore further the possibility that S-V bond formation can take place, another scenario has been explored, in which the source of proton might be the -SH group itself.

This would imply a more favorable (S⁻)-V bond formation. Indeed, even though the pH conditions (close to neutrality) at which Guedes da Silva et al conducted their experiments should not allow the formation of deprotonated thiolate groups, it is not fair to exclude a priori that one of the O atoms of amavadin could act as a base. Another possibility might entail the transient formation of a

deprotonated sulfur-protonated amavadin (transition) state with concerted (S⁻)-V bond formation.

The structures of the anionic intermediate identified are sketched in Scheme 2.



Scheme 2. The complete set of negatively charged adducts between MMA and $[V(\text{hidpa})_2]^-$ is shown. Relative stabilities are indicated in

black, in Å), delocalization indexes (grey) and atomic net charges (grey, italics). In blue is evidenced the MMA⁻ structure.

In Figure 5 are reported the optimized geometry and QTAIM parameters for the most stable anionic adduct. The V-S bond distance is 2.413 Å, which is on average 0.4 Å longer with respect to V-O, but the $\delta(\text{V-S})$ index is 0.811 electron pair, which is roughly 0.1-0.2 electron pairs higher than average $\delta(\text{V-O})$ values compute for oxygen atoms that belong to carboxylate group. To discuss the atomic net charges of this species, we need to consider again the V and S charges in amavadin and MMA. To estimate the redox state of V^N, we used a simple linear model

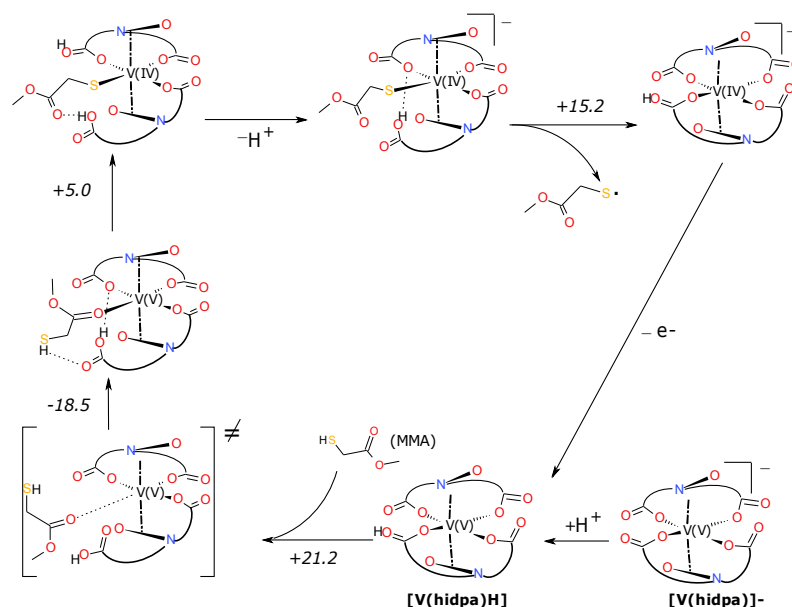
$$\Omega(\text{V})=0.12 \cdot \text{N}+1.402 \quad (4)$$

obtained by interpolation of the $\Omega(\text{V})$ values in V^{IV} and V^V amavadin. The V net charge in the anionic adduct is 1.901, which is less positive than that in V^V amavadin by 0.102 electron and corresponds to the V^{4.3} redox state according to the linear model. To properly discuss this result, we compute the QTAIM charges for the anionic form of the MMA. The $\Omega(\text{S})$ value of -0.785 electrons testifies to the localization of the negative charge in MMA⁻ on the sulfur atom, as one could expect. When MMA⁻ is coordinate through sulfur to V^V of amavadin, the S atom results less negatively charged by 0.469 electron, with a $\Omega(\text{S})$ value of -0.316. The two results presented indicate that upon MMA⁻ coordination i) the coordinated S atom is partially oxidized and conversely V is partially reduced; ii) a fraction of the S electron

density is shifted toward the V-S region to form the V-S bond. The remaining sulfur electron density is essentially delocalized on the oxygen atom of the V coordination, as showed by the average increase of their negative charges. Finally, it is interesting to note that according to the model (4), the V^V redox state in the most stable neutral intermediate does not change

5. Thiol oxidation: mechanistic considerations

The DFT characterization of reaction intermediates formed when MMA reacts with $[\text{V}(\text{hidpa})_2]^-$ allowed us also to shed some light on the catalytic mechanism (Scheme 3). According to our results, only the protonated V(V) form of amavadin can bind MMA, forming an intermediate species in which the carbonyl MMA group acts as a vanadium ligand. The latter observation is in agreement with the experimental observations indicating the necessity of a carbonyl moiety for turnover. The lowest energy MMA- $[\text{V}(\text{hidpa})_2\text{H}]$ isomer is in equilibrium with an isomer featuring direct coordination of the S atom of MMA to vanadium. Notably, the formation of the latter intermediate species corresponds to the redox step of the mechanism, in which the oxidation state of vanadium changes from V to IV and the MMA molecule is oxidized. The catalytic cycle for monoelectron oxidation of MMA is closed by release of the thiol radical and oxidation of the catalysts. Radical coupling of two oxidized MMA molecule leads to S-S bond formation.



Scheme 3. Proposed catalytic cycle for the reaction: $\text{MMA} \rightarrow \text{MMA}\cdot + \text{H}^+ + \text{e}^-$, as turnover by the protonated form of amavadin. Energies are indicated in $\text{kcal}\cdot\text{mol}^{-1}$. Those steps corresponding to protonation/deprotonation reactions and oxidation reactions are not labelled with energetic tags since by presently available DFT theory levels, both pKa and reduction potential calculations are affected with too large margins of error to be directly compared with experimentally observed values.

4. Conclusions

The present contribution shed light on energetic and structural features of an adduct which forms between amavadin and a prototypical substrate (MMA) in the thiol oxidation to the corresponding disulfide, a reaction of extreme importance in biological systems. In particular,

DFT shows that both thermodynamics (slightly endoergonic reaction) and kinetics (non-prohibitive reaction barrier) of the adduct formation are in line with an accessible process. QTAIM and structural parameters are consistent with the activation of the substrate when bound to the catalyst. Moreover, an indirect evidence on the role of the vanadium (in its d^0 active state) based catalyst protonation has been proposed.

Interestingly, theoretical results has allowed us to rationalize, on the basis of structure-function relation, one of the most relevant experimental observations by Guedes da Silva et al, which is here worth recalling: an ancillary carboxylate group (or its derivative ester function) is a crucial requisite in order to detect electrocatalytic activity. Thus, the identification by DFT of a species in which V^V is ligated by one of the ester oxygen atoms of MMA is strongly supportive of the role played in the whole oxidative process by this secondary functionality present on reactive substrates. Secondly, experimental data pointed out a precise geometrical requirement which had to be fulfilled by active substrates: ester/acid moiety must be two or three bonds away from the reducing $-SH$ center. Regard to this aspect, DFT results permit only speculative conclusions to be drawn. Actually, an *unstable* isomer adduct has been located (Scheme 1) which features a bis-chelate [O;S] coordination fashion of MMA to V, which implies a favorable intramolecular 5-membered ring formation. If methyl mercaptopropionate (another reactive substrate) were to replace MMA, one should conceivably assist to a 6-membered ring formation. According to a Michaelis-Menten catalysis scheme, after the substrate-catalyst adduct

has formed, its destabilization must occur to observe reaction acts. Such unstable form with chelate coordination of MMA might represent such reactive species.

Finally, some stereoelectronic features of the catalyst itself, already investigated in the past with less-sophisticated methodologies, have been revisited (and confirmed) by means of high-level theory calculations.

References

- 1) J. H. Trefy and S. Metz, *Nature*, 1989, **342**, 531.
- 2) D. Rehder, *Org. Biomol. Chem.*, 2008, **6**, 957.
- 3) E. Bayer and H. Kneifel, *Z. Naturforsch B*, 1972, 207.
- 4) H.-U. Meisch *et al.*, *Naturwissenschaften*, 1979, **66**, 620.
- 5) M. A. A. F. de C. T. Carrondo *et al.*, *J. Chem. Soc., Chem. Commun.*, 1988, 1158.
- 6) E. M. Armstrong *et al.*, *J. Am. Chem. Soc.*, 1993, **115**, 807.
- 7) H. Kneifel and E. Bayer, *Angew. Chem., Int. Ed. Engl.*, 1973, **12**, 508.
- 8) C.M.M. Matoso *et al.*, *ACS Symp. Ser.*, 1998, **711**, 241.
- 9) P. M. Reis, *et al.*, *Chem. Commun.*, 2000, 1845.
- 10) M. V. Kirillova *et al.* *J. Am. Chem. Soc.*, 2007, **129**, 10531.
- 11) M. V. Kirillova *et al.*, *Chem.–Eur. J.*, 2008, **14**, 1828–1842.
- 12) T. Hubretsge *et al.*, *Eur. J. Org. Chem.*, 2007, 2413–2422.
- 13) R. D. Thackerey and T. L. Riechel, *J. Electroanal. Chem.*, 1988, **245**, 131.
- 14) M. F. C. G. da Silva *et al.*, *J. Am. Chem. Soc.*, 1996, **118**, 7568.
- 15) E. M. Armstrong *et al.*, *J. Chem. Soc., Dalton Trans.*, 1995, 191.
- 16) C. Remenyi *et al.*, *J. Phys. Chem. B*, 2005, **109**, 4227.
- 17) K. J. Ooms *et al.*, *Dalton Trans.*, 2009, (17), 3262–3269.
- 18) K. R. Geethalakshmi *et al.*, *Inorg. Chem.*, 2007, **46**, 11297.

- 19) D. C. Crans *et al.*, *Inorg. Chem.*, 2010, **49**, 4245.
- 20) A. D. Becke, *Phys. Rev. A: At., Mol., Opt. Phys.*, 1988, **38**, 3098–3100; J. P. Perdew, *Phys. Rev. B.*, 1986, **33**, 8822.
- 21) K. Eichkorn *et al.*, *Theor. Chem. Acc.*, 1997, **97**, 119.
- 22) R. Ahlrichs *et al.*, *Chem. Phys. Lett.*, 1989, **62**, 165.
- 23) A. Schafer *et al.*, *J. Chem. Phys.*, 1994, **100**, 5829.
- 24) G. Zampella *et al.*, *Inorg. Chem.*, 2006, **45**, 7133.
- 25) G. Zampella, *et al.*, *J. Am. Chem. Soc.*, 2005, **127**, 953.
- 26) A. Klamt, *J. Phys. Chem. A*, 1995, **99**, 2224.
- 27) A. Klamt, *J. Phys. Chem. A*, 1996, **100**, 3349.
- 28) A. Klamt and G. Schüürmann, *J. Chem. Soc. Perkin Trans.*, 1993, **2**, 799.
- 29) F.W. Beigler-Konig *et al.*, *J. Comput. Chem.*, 1982, **317**, 3.
- 30) C. Gatti, SF_ESI Code, private communication.
- 31) X. Fradera *et al.*, *J. Phys. Chem. A.*, 1999, **103**, 304.
- 32) P. Macchi and A. Sironi, *Coord. Chem. Rev.*, 2003, **238–239**, 383.
- 33) C. Gatti and D. Lasi, *Faraday Discuss.*, 2007, **135**, 55.
- 34) L. Bertini *et al.*, *Organometallics*, 2010, **29**, 2013–2025.
- 35) R. E. Berry, *Angew. Chem. Int. Ed.*, 1999, **38**, 795.
- 36) C. Lee *et al.*, *Phys. Rev. B*, 1988, **37**, 785.
- 37) S. Vosko *et al.*, *Can. J. Phys.*, 1980, **58**, 1200.
- 38) A critical point is a stationary point of the electron density characterized by the rank (the sum of the non-zero hessian eigenvalues) and the signature (the sum of the sign of the element of the diagonalized hessian). A bond critical point (BCP) denotes a (3,-1) first-order saddle point of electron density along the maximum electron density line between two atoms.