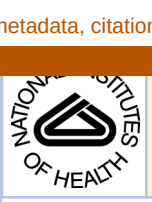

Published in final edited form as:
Theor Chem Acc. 2007 May 1; 117(5-6): 723–741. doi:10.1007/s00214-006-0203-4.

Computational approaches to shed light on molecular mechanisms in biological processes

Giorgio Moro,
Dipartimento di Biotecnologie e Bioscienze, Università di Milano-Bicocca, Piazza della Scienza, Milano 20126, Italy

Laura Bonati,
Dipartimento di Scienze dell'Ambiente e del Territorio, Università di Milano-Bicocca, Piazza della Scienza, Milano 20126, Italy

Maurizio Bruschi,
Dipartimento di Scienze dell'Ambiente e del Territorio, Università di Milano-Bicocca, Piazza della Scienza, Milano 20126, Italy

Ugo Cosentino,
Dipartimento di Scienze dell'Ambiente e del Territorio, Università di Milano-Bicocca, Piazza della Scienza, Milano 20126, Italy

Luca De Gioia,
Dipartimento di Biotecnologie e Bioscienze, Università di Milano-Bicocca, Piazza della Scienza, Milano 20126, Italy

Pier Carlo Fantucci,
Dipartimento di Biotecnologie e Bioscienze, Università di Milano-Bicocca, Piazza della Scienza, Milano 20126, Italy

Alessandro Pandini,
Dipartimento di Scienze dell'Ambiente e del Territorio, Università di Milano-Bicocca, Piazza della Scienza, Milano 20126, Italy

Elena Papaleo,
Dipartimento di Biotecnologie e Bioscienze, Università di Milano-Bicocca, Piazza della Scienza, Milano 20126, Italy

Demetrio Pitea,
Dipartimento di Scienze dell'Ambiente e del Territorio, Università di Milano-Bicocca, Piazza della Scienza, Milano 20126, Italy

Gloria A.A. Saracino, and
Dipartimento di Biotecnologie e Bioscienze, Università di Milano-Bicocca, Piazza della Scienza, Milano 20126, Italy

Giuseppe Zampella
Dipartimento di Biotecnologie e Bioscienze, Università di Milano-Bicocca, Piazza della Scienza, Milano 20126, Italy

Giorgio Moro: giorgio.moro@unimib.it

Abstract

Correspondence to: Giorgio Moro, giorgio.moro@unimib.it.

Computational approaches based on Molecular Dynamics simulations, Quantum Mechanical methods and 3D Quantitative Structure-Activity Relationships were employed by computational chemistry groups at the University of Milano-Bicocca to study biological processes at the molecular level. The paper reports the methodologies adopted and the results obtained on Aryl hydrocarbon Receptor and homologous PAS proteins mechanisms, the properties of prion protein peptides, the reaction pathway of hydrogenase and peroxidase enzymes and the defibrillogenic activity of tetracyclines.

Keywords

Prion protein peptide; 3D-QSAR; Protein structure prediction; Protein structural flexibility; DFT calculations; Catalysis

1 Introduction

The computational approach taken in our research on biological processes focuses mainly on three methodological areas. One includes a variety of methods based on Molecular Mechanics (MM), Molecular Dynamics (MD) and Statistical Mechanics, and Monte Carlo methods (MC) [1–4]. The second is an approach based on advanced Quantum Mechanical (QM) methods applied to model systems, the aim being to obtain an accurate description of the enzyme reaction path [5–9]. The third is an approach aimed at obtaining statistical models through an analysis of data inferring relative Quantitative Structure-Activity Relationships (QSAR) [10–17].

As is well known, approaches based on MD and MC theories are the only ones presently available to study complex systems like proteins in solution; in such approaches all water molecules are explicitly considered to be contained in a well dimensioned, and shaped (periodic), box. Thus the approach to the problem of protein structure at the classical level is even more acute when there is the modelling of interaction between proteins themselves, between protein and DNA fragments or between protein and substrates (as in drug discovery, toxicology studies or virtual enzyme engineering).

Indeed, MD or MC methods based on classical interaction potential are orders of magnitude faster than corresponding QM methods which, presently, cannot be applied for an accurate description to whole protein systems, but only to a small part of the entire system.

However, MD and MC methods are not completely free of difficulties, which are generated just by the very high number of degrees of freedom (about 10^5). In practice it is impossible to sample the phase space exhaustively due to the limitations in reliability of the final results, which are seen in the following: (a) limited sampling (time of simulation), this necessarily corresponds to a relatively short time window during which only specific conformational transitions can occur, these being characterized by very low activation energy or high frequency; (b) low frequency motions and geometry rearrangements not usually accessible within the normal simulation times (10–30 ns); this is a particularly severe drawback as, from the biological point of view (e.g. dynamics of folding/unfolding), long term structural evolutions are very relevant; (c) standard time step strategies do not give a reliable estimate of the true frequency of conformational transitions, with the obvious consequence that also the energy barriers cannot be quantitatively estimated.

Given our awareness of the difficulties involved, we took great care when applying the MD and MC methods (discussed in the following sections) to maximize the degree of phase space sampling, using the repeated trajectory technique starting from the same values of

potential energy and total kinetic energy but with different distributions of the atomic velocity components. Furthermore, we adopted the *essential dynamics* [18] technique extensively, in order to extract the low frequency motions of biological relevance. The projection of a whole trajectory, or of selected trajectory time windows over a subspace spanned by a few eigenvectors of the covariance matrix, allows a clearer interpretation of the dynamic behaviour of the biomolecules. Specific examples of such an analysis will be presented in the following sections. An additional aspect, connected to the actively investigated protein structure determination, is a new computational scheme which should produce a reliable score function for the quality of the protein structure derived from homology modelling procedures coupled with MD refinement. Several energy functions (*score functions*) have been proposed to classify protein structure quality, but all the proposals make use of global information of the whole structure. In this context, small deviations localized in very specific protein regions lose their importance. This could be quite a serious drawback as local features can be determinants for the functional properties of a considered protein. The aim of the investigation described below was to obtain score functions, based on global and local features, that make it possible to discriminate between *correct* and *wrong* structural models.

In the area of modelling reaction pathways using QM methods, active research is in progress to unravel structure–activity relationships in metal enzymes which control very basic processes, like the hydrogenases (iron or nickel–iron enzymes promoting reversible reduction-oxidation of hydrogen) and vanadium peroxidases (enzymes promoting a variety of oxidation reactions).

In the field of QM modelling for transition metal complexes, DFT methodology has triggered extensive modelling, leading to a very consistent increase in the area of applicability together with a consistent reduction in the computer time needed. Development has been such that we can now employ a large basis set (TZVP) in our studies, and feel quite confident about the theoretical results. As for the computed quantities, the level of theory adopted guarantees the evaluation, at a quantitative level, of best molecular geometries, reactions and activation free energies, as well as of vibrational frequencies and associated atomic displacements.

However, the QM investigation of metal enzymes is limited by the fact that such complex systems have to be simplified via the introduction of relatively small molecular models, which, while keeping all the chemical features of the active site, necessarily ignore the long range effects caused by the protein environment. Such limitations could be (partly) overcome using a QM/MM mixed approach [19].

In this procedure only atoms of the enzyme active site explicitly involved in reaction with substrates are treated using quantum chemistry, whereas the chemical groups surrounding the active site, but not interacting directly with the substrate, are described classically, i.e. using MM force-fields. However, it should be pointed out that, even though the QM/MM approach has already been successfully employed to investigate the chemistry of enzymatic systems [20–22], some theoretical issues still exist, especially those related to the description of the boundary between the QM and MM regions.

In the area of the statistical modelling of quantitative structure-activity data (QSAR), new methods aimed at solving the problems of feature selection and model derivation are being searched for. One problem lies in the fact that the number of possible molecular descriptors on which a QSAR model could be based is usually extremely large, and certainly has redundant descriptors with respect to the *essential* subspace. In addition, very large descriptor space dimensions are also usually accompanied by the appearance of noise and

biasing. Different, well-suited, strategies have been proposed [23,24] to select only those descriptors likely to yield regression models with predictive ability. Recently, we developed a local variable reduction technique [25], and applied it in our laboratories to reduce the number of variables to the minimum while keeping the information content to the maximum. Moreover, the use of genetic algorithms allows the introduction of general flexibility in the search of the best model and, at the same time, offers a simple solution to model validation.

The present paper presents our recent results, the work presently in progress and the planned future development of computational studies on biological systems. The main focus is on the following subjects:

- Homology modelling and Molecular Dynamics to investigate the activation mechanism of the Aryl hydrocarbon Receptor and homologous PAS proteins;
- Structure-function relationships in specific protein families: towards the combined use of Molecular Dynamics and structural prediction method;
- Molecular Dynamics simulations as a tool for the investigation of structural flexibility and aggregation propensity in prion protein peptides;
- Quantum mechanical methods in Bioinorganic Chemistry: the activation of small molecules in hydrogenase and peroxidase enzyme pathways;
- Developing and applying suitable strategies for variable reduction in 3D-QSAR investigation: the defibrillogenic activity of tetracyclines.

Each subject is briefly introduced, and the results are discussed in Sect. 3.

2 Methods and computational details

2.1 Homology modelling

The templates were structurally aligned pairwise according to DALI [26]. The sequence alignments were obtained by CLUSTALW [27] and the results were also confirmed by the Align-2D option within MODELLER 8v1 [28]. The protein three-dimensional models were generated by MODELLER 8 v1, and model quality was assessed by the PROCHECK [29] and PROSAII [30] programs. Three-dimensional visualizations were generated by PyMOL [31].

2.2 Threading

Different threading methods (sequence-structure fitness) were used to detect remote similarities with proteins of known 3D structure: 3D-PSSM [32], MGen-THREADER [33], 123D+ [34], Fugue [35] and Topits [36]. Three-dimensional models were built using the Jackal protein modelling software package (http://tran-tor.bioc.columbia.edu/~xiang/jackal). Target-template alignment was submitted to the sub-program Nest, which generates a 3D model on the basis of a given alignment; the program optimizes the geometry in torsional space to remove clashes between atoms, and finally optimizes the loop regions that are characterized by the presence of gaps in the alignment. As the final step, the models underwent molecular mechanics optimization.

2.3 Molecular dynamics simulations

All the simulations were performed with GROMACS 3.1.4 [37,38] using the GROMOS96 43a2 version of the GROMOS force field; where possible, the version implemented on parallel architecture was used.

The ionization state of charged residues in proteins was set to mimic a neutral pH environment: Lys and Arg residues were positively charged, whereas Asp and Glu were negatively charged. The protonation state of the histidine residues was predicted with GROMACS tools and confirmed by visual inspection of the molecular environment of each histidine. In order to make the overall systems neutral, a number of water molecules equal to the net charge of the protein was replaced by ions of opposite charge (Cl^- and Na^+). For peptide systems, the effect of varying the pH was mimicked by changing the protonation state of the His residues.

Protein structures, including crystallographic water molecules when present, were soaked in cubic or dodecahedral boxes containing SPC water molecules [39]. In the case of simulation on peptide systems, each peptide (amidated at the C-terminal and acetylated at the N-terminal) was simulated as an 'isolated molecule' in four solvents of different polarity: water, dimethyl-sulfoxide (DMSO), hexane and trifluoroethanol (TFE).

The simulations were performed in the NPT or NVT ensembles, applying periodic boundary conditions. Protein and solvent were coupled independently with a 300 K thermal bath using the Berendsen thermostat and a coupling period of 0.1 ps [40]. Long-range electrostatic interactions were calculated using the Particle-mesh Ewald (PME) summation scheme [41].

The internal degrees of freedom of water molecules were constrained by the SHAKE algorithm, while all bond distances in the proteins were constrained using the LINCS algorithm [42]. The time step was set to 2 fs and, for each system under study, several (4–10) simulation replicas (10–60 ns trajectories) were generated starting with different initial configurations, initial velocity and box dimension. For specific simulations the interacting site method was employed to allow a 4 fs time step.

In some studies, simulation efficiency and reliability were evaluated by RMSD matrices, overlapping of replicas on combined trajectories, and cosine content of the first eigendirections. Moreover, Essential Dynamics analysis [18], performed on the C_α atoms of selected residues, was applied to MD trajectories to extract informative directions of motion.

2.4 Quantum mechanical calculations

Theoretical investigations of models of the investigated active site were mainly carried out within the Density Functional Theory (DFT) framework, as implemented in the TURBOMOLE program [43], using the hybrid three parameter B3LYP and pure BP86 functionals in conjunction with a valence triple ζ basis set with polarization on all atoms (TZVP) [44]. A restricted framework was employed for closed-shell electronic structures, whereas unrestricted Kohn-Sham calculations were carried out for open-shell complexes. Stationary points of the energy hypersurface were located by means of energy gradient techniques, and full vibrational analyses were carried out to further characterize each stationary point.

Free energy (G) values were obtained from electronic SCF energy considering three contributions to the total partition function (Q), namely $q_{translational}$, $q_{rotational}$, $q_{vibrational}$, under the assumption that Q may be written as the product of such terms [45]. In order to evaluate enthalpy and entropy contributions, the temperature and pressure values were set to 298.15 K and 1 bar, respectively. Rotations were treated classically, and vibrational modes described according to harmonic approximation. Effects due to proteic environment were simply modelled according to the COSMO approach [46,47]. To better evaluate environmental effects, the calculations were carried out in vacuum (ε = 1), and considering a polarizable continuum medium characterized by two different ε values (ε = 4 and 40) commonly used to model proteins [48–52].

2.5 QSAR modelling

All the compounds were considered in their neutral protonation state and in the zwitterionic form. Partial atomic charges were calculated by fitting the MEP computed at the *ab initio* HF-SCF/6-31G** level. The lowest energy conformations of the compounds were aligned by rigid-body superimposition of common structural moiety. A grid of 32× 29× 29 Å dimension and 0.5 Å spaced points was built, embedding all the molecules and discarding all the points inside the envelope of the van der Waals surface of at least one molecule. At each grid point Molecular Electrostatic Potential (MEP) and non-bonded (van der Waals, vdW) interaction energies were calculated for all the molecules, using the potential derived atomic charges and TRIPOS force field [53] parameters.

At each point, the Pearson correlation coefficient values between biological activity and MEP (R_{MEP}), as well as vdW (R_{vdW}), were calculated. Then, points corresponding to local maximum or minimum values of R, with respect to the first and the second nearest points, were searched for. Only local maxima or minima points, with $R > |0.2|$ and Standard Deviation of MEP and vdW >1.0 kcal mol^{-1} and >0.05 kcal mol^{-1}, were retained. Starting from about 200,000 grid points, the variable reduction procedure provided 33 (R_{MEP}) and 140 (R_{vdW}) variables to be used in the search for the most predictive models.

Genetic Algorithm-Variable Subset Selection (GA-VSS) method [54,55] was used to search for the best least squares regression models, optimizing the model prediction power expressed by the leave-one-out cross validated regression coefficient Q^2_{LOO}. The search was restricted to models containing from 4 to 6 independent variables and several GA-VSS runs have been performed, each starting from a random population of 100 individuals. Finally, a fully cross-validated regression coefficient, $Q^2_{20\%}$, with 20% of cancellation groups, was calculated for the most effective models.

3 Results and discussion

3.1 Homology modelling and molecular dynamics to investigate the activation mechanism of the Aryl hydrocarbon receptor and homologous PAS proteins

Besides the number of computational studies carried out in the last decade for the development of new drugs or chemicals, there has arisen, more recently, a need to further develop methodologies for the prediction of more complex phenomena such as toxicity. The manifestation of a toxic phenomenon is the result of a cascade of biochemical events and transformations that, in the cases of receptor-mediated toxicity, involve high affinity binding to a receptor as the initial step. In several cases structural information on the receptor is lacking, thus the combined use of bioinformatics and molecular modelling methods can prove very effective.

In this framework we have, in recent years, performed a study on molecular events that regulate the activation mechanism of the Aryl hydrocarbon Receptor (AhR). This is a cytoplasmatic protein that plays a key role in mediating the toxic and biological effects of a variety of chemicals, among which are environmental contaminants such as the halogenated aromatic hydrocarbons (HAHs) and the polycyclic aromatic hydrocarbons (PAHs). Exposure to numerous HAHs, including 2,3,7,8-TetraChloro Dibenzo-p-Dioxin (TCDD), the most potent member of this class of chemicals, produces a wide variety of species- and tissue-specific toxic and biological effects [56]. In the cytosol, AhR is present as a multiprotein complex that, following ligand binding, is presumed to undergo conformational changes, exposing a nuclear localization sequence, and then to translocate into the nucleus. Here AhR dissociates from the protein complex and dimerizes with ARNT (AhR Nuclear

Translocator). Finally, the liganded heterodimer can interact with its specific DNA recognition site [57].

The aim of our work was to model ligand-receptor interactions at the molecular level to rationalize the differences observed in ligand binding affinities, as well as to highlight conformational changes in the ligand binding domain (LBD) that lead to AhR activation and the signaling mechanism. In the past such an analysis was prevented by the lack of information on the three-dimensional structure of the AhR LBD. However, in recent years, the availability of the experimental structures of homologous proteins belonging to the Per-ARNT-Sim (PAS) superfamily has allowed us to develop a homology model of mouse AhR LBD, and to propose initial hypotheses regarding the domain features important to its functional activity. Site-directed mutagenesis experiments have also been performed, aimed at validating our proposal by analysing the effects of mutations in some key positions on the AhR TCDD-and DNA-binding.

Comparative modelling work suggested the possibility of obtaining information on the signal transduction mechanism of AhR, by comparing it with the known mechanisms of homologous proteins belonging to the PAS superfamily. These proteins cover a wide spectrum of cellular responses, with highly specialized mechanisms, to different signals. Despite the evolutionary divergence of the sequences, these proteins share a high structural conservation of the receptive domain. Indeed, in a study on the dynamical behaviour of these domains, Molecular Dynamics methods were applied to them in order to obtain comparative information on their conformational and functional properties.

The first theoretical model for mAhR LBD was developed [58,59] by homology modelling techniques, on the basis of the crystal structures of three homologous PAS proteins: the bacterial photoactive yellow protein, PYP, the human potassium channel HERG, and the heme binding domain of the bacterial O_2 sensing FixL protein. Despite the low level of sequence identity, the three PAS structures show a high structural conservation of the α+β fold. This includes a five-stranded β-sheet hung on a long central helix (helical connector) and a bulge of three small helices. At that time the analysis suggested FixL as the best template structure [60]. Despite the low atomic resolution, due to the low sequence identities with the available reference structures, this first structural proposal (Fig. 1a) led to preliminary hypotheses on the residues that might be involved in ligand binding, and these stimulated site-directed mutagenesis experiments. In a study aimed at demonstrating whether AhR can mediate enhanced transcriptional activity in the absence of ligand binding [61], two point mutants, A375I and A375F, were generated on the basis of the above model. In fact, the observation that the ALA375 side chain protrudes into the middle of the modelled lig-and binding pocket suggested that the introduction of side chains with considerably increased steric hindrance in such a key position could hinder ligand binding. Both mutants failed to bind ligands or exhibit enhanced activity in cells exposed to AhR ligands, thus confirming our preliminary hypothesis on the ligand binding pocket location and general features.

A number of PAS structures have been derived by NMR spectroscopy or X-ray diffraction in recent years: that of the fern photoreceptor, Phy3, of the human PAS Kinase, hPASK, of the human hypoxia-inducible factor 2α, HIF-2α, of the mouse nuclear coactivator 1, NcoA, of the *Drosophila* clock protein PERIOD, dPER and of the *human* Aryl Hydrocarbon Receptor Nuclear Translocator, ARNT. Among these structures, those of HIF-2α [62] and ARNT [63] deserve particular attention as these proteins share some key functional similarities with AhR and, accordingly, their PAS B domains show the highest degrees of sequence identity (31.1 and 21.2%) and similarity (62.1 and 53.8 %) to mAhR, considering all the presently available PAS structures.

In the different PAS domains the helical connector is displaced from the β-sheet in different ways, designing a cavity suitable for arranging different kinds of cofactors. Therefore the choice of the template structures is crucial for modelling the AhR LBD and, consequently, the ligands' access to the cavity. To address this problem, different mono- and multi-template models were developed and the quality of modelled structures was carefully assessed [64,65]. On the basis of this analysis, the bi-template model based on HIF-2α and ARNT PAS domain structures (Fig. 1b) was selected as the most reliable one. As expected, while the general features of the different modelled LBDs are conserved, differences are observed in the helical connector position and, consequently, in the cavity entrance location. In the most reliable models the cavity entrance resembles those of HERG, Phy3, HIF-2α and ARNT, and the cavity entrance falls at the opposite side of the helical connector with respect to FixL and the first mAhR model (compare Fig. 1a and b).

To confirm the features of the modelled binding cavity, site-directed mutagenesis experiments were performed [64,65]. The effects of point-mutations were measured on the basis of TCDD- and DNA-binding activity. The results support the proposed HIF-2α/ARNT model, and confirm that the modelled cavity is indeed involved in ligand binding. In fact, mutations of residues that fall beyond the modelled LBD pocket do not affect AhR TCDD- and DNA-binding, whereas residues for which there is at least one mutation adversely affects the binding point in the modelled cavity. Some of the mutagenesis results are also consistent with the binding cavity features of the previous model. For example, mutations of ALA375, which in both models fall in the middle of the LBD pocket, indicate more hindering residues, dramatically reducing the TCDD and DNA binding of mAhR. However, three of the residues, in which mutation does not affect binding and that accordingly reside on the external domain surface in the HIF-2α/ARNT model, lie in key positions inside the pocket in the FixL-based model [58,59]. These observations support our hypothesis concerning the greater reliability of the last model and the proposal of ligand access to the interior of the binding cavity.

In recent years, significant efforts have also been directed towards rationalising the mechanisms employed by PAS domains to convert input stimuli into signals that propagate to downstream partners. One of the most intriguing suggestions that has emerged is that the functionality of the PAS module is intrinsically dynamic, and that flexibility at the receptive sites plays a central role in promoting significant pathways of conformational changes, subsequently transmitted to extra-domain units through suitable domain interfaces. Therefore to gain insight into the dynamic features of the superfamily, a comparative analysis was performed [66] on the conformational flexibilities of a set of known PAS structures. These include the X-ray structures for the receptive states of PYP, HERG, FixL and Phy3, as well as the NMR ensembles of structures of hPASK and HIF-2α.

The computational cost of nanosecond Molecular Dynamics simulations is quite high, especially when the use of explicit solvent and detailed long-range electrostatic descriptions are necessary to correctly address the physical behaviour of the system. However, recent work has reported that incorrect sampling can be misleading in analysing molecular dynamics results, especially when studies involve Essential Dynamics where insufficient sampling can lead to a complete misinterpretation of the results, masking diffusive motion in random directions within the shape of essential motion [67,68]. Effective indexes of sampling reliability are the extension of conformation re-sampling in the phase space, evaluated by the RMSD matrices, the overlap values of increasing portions of the simulation with the overall trajectory [68] and the cosine content of the first eigen-directions derived from the covariance analysis [67]. On the basis of the evaluation of these indexes it was concluded that extensive conformational space sampling is needed to obtain a reliable picture of the dynamics of PAS structures. In fact, while none of these was extensively

described by a single 10 ns simulation, the choice of joining four 10 ns simulation replicas, obtained by starting from the same structure but with different sets of atomic velocities, guaranteed a better sampling with acceptable setting up and computational times.

Essential Dynamics analysis is a well documented application of Principal Component Analysis to Molecular Dynamics data [18], aimed at extracting informative directions of motion in a multidimensional space, thus reducing the overall complexity of the simulation and isolating motion important to the system. The advantage of this analysis is that it gives the possibility of highlighting spatially correlated motions, and of extracting most of the useful information located in the low frequency modes that are usually involved in conformational transitions. Essential Dynamics analysis was successfully employed to extract the relevant information from the sampled conformational spaces. A neat separation of the motion across the re-oriented space was observed, defining a reduction in the space to be analysed to 6 principal directions. This is compatible with the inclusion of more than 50% motion in the subspace, all directions being employed with a significant amount of variance on displacement, and the inclusion of all anharmonic modes of motion.

Analysis of the RMSF with respect to the residue position in the Essential Subspace allowed the definition of regions with higher mobility within each domain (Fig. 2). The location of the obtained “motion patterns” was in complete agreement with available experimental evidence.

Comparison of motion patterns in the essential subspaces was performed by taking, as reference, the structural alignment of the domains obtained by DALI. The highest differentiation in the dynamic behaviour of the PAS domains was found in the highly flexible helical zone. Both the structural comparisons and the dynamical analysis suggest that this central part of the domain contains the specialized segment of the sequence, where the receptive function is shifted to different parts through proper loop length adjustment and related changes in the dynamical properties. Interestingly, this specialization appears to have evolved in the different domains by following a specific trend that highlights the occurrence of three general solutions developed to sense external signals. Moreover, dynamic and structural features consistent with the above three solutions were found to be conserved also in domains not containing cofactors; in these cases, the role of triggering signal transmission was assumed by external ligands that adopted one of these paths on entering the binding cavity.

As both experiments [69] and ED analysis evidenced, the PYP light reception mechanism involves, in the first stages, conformational changes in the chromophore hydrogen-bond network at the N-terminal cap and, in the following, signal propagation from the N-terminal helices to extra-domain helices that are in close contact. The similar location of significant flexibility in PYP and HERG suggests for the latter the involvement of this part of the structure both in signal reception and propagation. On the other hand, the observed conservation of motion patterns in hPASK and FixL confirms the common involvement of the long loop between the helical connector and the β-sheet in both ligand binding and interdomain interactions, that was also evidenced by NMR-based studies [70]. Interestingly, the detected similarities between the Phy3 and the HIF-2α motion patterns (Fig. 2), suggest that the latter could exploit its activity by reception external signals in the cleft between the N-terminal cap and the helical connector, where FMN exploits receptor activity in Phy3 [71]. Although a natural ligand has not been identified for HIF-2α, based on the parallels with other PAS signaling mechanisms (including that of the highly similar AhR), it was hypothesized that small organic compounds can bind the HIF-2α PAS domain and generate structural changes that are then transmitted to the β-sheet, identified as the interface of heterodimerization. In the model derived by the flexibility analysis, ligands are proposed to

enter the binding cavity, following the outlined path, as a consequence of the intrinsic flexibility of the helical connector and its hinge loops, that dynamically open the gate to the internal pocket.

The agreement between the homology modelling results and the site-directed mutagenesis experiments, along with the higher resolution of the mAhR LBD model derived by the HIF-2 α and ARNT templates, provides a reliable framework for developing further hypotheses and for addressing more specific molecular modelling studies. These will be mainly focused on comparisons among AhRs of different species, on ligand-receptor interactions, and on the characterization of the protein-protein interactions. Moreover, once annotation and comparison of the dynamic properties is extended to newly available PAS structures, it will be possible to employ the model of receptive activity specialization presented here in order to gain insight into AhR activation mechanisms.

Finally, the PAS superfamily analysis demonstrated how comparison of functional motions among distant homologous proteins with conserved fold characteristics could give some insight into their functional specialization. On the basis of these results, a procedure to perform a large-scale comparison of protein flexibilities was designed and assessed [72]. The use of a fast conformational sampling method and a simple and synthetic index of similarity between domain flexibilities, calculated on the residue base, have been proposed. The possibility of identifying family relationships on the basis of the dynamical features has been verified for a group of protein domains belonging to an ASTRAL/SCOP40 fold.

3.2 Towards the combined use of molecular dynamics and structural prediction methods to unravel structure-function relationships in specific protein families

Computational methods capable of helping infer structure-function relationships in proteins are, nowadays, key tools in chemical-biology due to the large and ever increasing amount of data obtained by genome sequencing. In fact, the efforts made, and the progress achieved, in high throughput X-ray and NMR methods are complemented by the development and critical application of computational methods suitable to infer three-dimensional properties of biological macromolecules. Prompted by these observations, we recently undertook studies aimed at the prediction of the structural and functional properties of some proteins, using computational methods such as homology modelling, fold recognition and molecular dynamics simulations. Three examples, concerning trypsins, histones and guanine nucleotide exchange factors, are briefly outlined in the following.

Trypsin-like serine proteases invariantly bind a calcium ion, which has been proposed to be involved in protein structure stabilization against thermal and proteolytic degradation, such as autolysis phenomena [73]. However, the molecular details related to the role played by the metal ion are still largely unknown. Therefore, we carried out molecular dynamics simulations to investigate the dynamic behaviour of bovine and salmon trypsins in calcium-bound and calcium-free forms, with the aim of evaluating the role of the calcium ion in trypsin three-dimensional structure and autoproteolysis propensity. It turned out that calcium-free trypsins are characterized by a more flexible structure, revealing the molecular connection between Ca^{2+} binding and autoproteolysis propensity. Our results revealed that the removal of Ca^{2+} increases the flexibility of regions around its binding site, and also leads to the channeling of fluctuations to remote sites of the protein, possibly involving the interdomain loop. Moreover, it was shown that two primary autolysis sites are strongly influenced by calcium binding in bovine trypsin, whereas Ca^{2+} plays a less crucial role in salmon trypsin [74].

We also recently investigated, by fold recognition methods and molecular dynamics, histones, which are short proteins involved in chromatin packaging. We focused our

attention on the double histone fold in which there assemble two consecutive regions characterized by the typical structure of histones, thus originating a histone pseudodimer. This fold is included in a few prokaryotic histones and in the regulatory region of guanine nucleotide exchange factors of the Sos family [75]. The application of several secondary structure predictions and fold recognition methods demonstrated that also the viral protein gi|22788712 is compatible with the structure of a histone pseudodimer. Further computational analyses revealed that this protein module could retain the ability to mediate protein-DNA interactions, and could consequently act as a DNA-binding domain, suggesting a possible functional role in viral pathogenicity for this novel double histone fold domain [76].

Another recent line of research was aimed at addressing structure-function relationships in $Cdc25^{Mm}$, which is a mammalian Ras-specific guanine nucleotide exchange factor (GEF). Using homology modelling we showed that $Cdc25^{Mm}$ shares with Sos-GEF the structure of the putative catalytic HI hairpin, where the dominant negative T1184E mutation is located. In conjunction with experimental results, our data showed that nucleotide re-entry and Ras/GEF dissociation require GEF regions different from the HI hairpin [77].

In parallel to the development and improvement of prediction methods, reliable and accurate evaluation tools are needed to check the quality of computational protein models. In this context, we are presently developing a computational tool in which several structural and physical parameters that can be computed on a protein structure are weighted by a neural network, with the aim of obtaining an empirical energy function suited to discriminate among correct and incorrect protein models [78].

3.3 Molecular dynamics simulations as tools for the investigation of structural flexibility and aggregation propensity in prion protein peptides

Prion diseases [79,80] are a large group of fatal neuro-degenerative disorders including Creutzfeldt-Jacob disease (CJD), Gerstmann–Straussler–Sheinker (GSS) syndrome and fatal familial insomnia (FFI) in humans, bovine spongiform encephalopathy (BSE) in cattle and scrapie in sheep. Their pathogenesis appears associated with a post-translational conversion of the prion protein from its cellular form (PrP^{C}) to an abnormal and infective form (PrP^{Sc}). The PrP^{Sc} form is characterized by a high propensity to aggregate in amyloid fibrils and a part-resistance to protease cleavage. Moreover, in accordance with the protein-only hypothesis [81,82], PrP^{Sc} acts as a template for the conformational modification of PrP^{C} units. It is still not clear whether toxicity is related to the deposition of fibrils or to the pre-fibrillar oligomers formed in the early steps of the aggregation, though there is increasing evidence in favor of the second hypothesis [83]. Spectroscopic studies (FTIR and CD) showed that the PrP^{C} to PrP^{Sc} conversion is characterized by a decrease in α-helix (from 40 to 30%) content and an increase in β-sheet content (from 3 to 43%) [84,85]. The three-dimensional structure of PrP^{C}, resolved for human and other animal species, presents a flexibly disordered N-terminal region (23–124 segment) and a structured globular domain (125–231 segment) [86–88]. The flexibility of the N-terminal could contribute to lower the interconversion barriers between different secondary structure elements involved in PrP^{C} to PrP^{Sc} transition. Because of its complex oligomeric nature, the PrP^{Sc} structure has still not been resolved, although some indications of its structural preferences have been obtained from FTIR [89], electron microscopy [90,91] and epitope mapping experiments [92,93].

To understand the mechanism underlying conformational transition and the aggregation process, several prion peptides have been studied. Even so, this apparently simplified approach retains objective experimental difficulties because of the highly aggregating propensity exhibited by such peptides. In these situations the use of theoretical methods could be helpful, and many theoretical studies concerning prion protein and peptides have

been conducted. Among such studies is also our work aimed at investigating the conformational flexibility of prion peptides using Molecular Dynamics techniques.

First of all, we focused our attention on the synthetic homologue of residues 106–126 of human PrP (PrP106-126). This peptide exhibits some properties typical of PrP^{Sc}. These include neurotoxicity, ability to activate astroglial and microglial cells [94–96] and a tendency to aggregate into amyloid fibrils that are partly resistant to protease digestion [97–99]. Moreover PrP106-126 shows remarkable polymorphism, acquiring different secondary structures in different environments, as evidenced by circular dichroism (CD) spectroscopy [100,101] and nuclear magnetic resonance (NMR) experiments [102,103]. These have shown that several chemical-physical conditions, such as pH, ionic strength and solvent composition influence the secondary structure of the peptide. Due to these properties the PrP106-126 peptide is a widely studied model system for the in vitro investigation of the pathological PrP protein. The PrP106-126 peptide consists of an N-terminal polar head (Lys-Thr-Asn-Met-Lys-His-Met) followed by a long hydrophobic tail (Ala-Gly-Ala-Ala-Ala-Ala-Gly-Ala-Val-Val-Gly-Gly-Leu-Gly). Nevertheless, the strategic position of segment 106–126 lying between the structured and unstructured domains of the prion protein means that this region can play a key role in the structural transition of PrP^{C} to PrP^{Sc}. PrP106-126 peptide [103,104] and the slightly different PrP portion, PrP109-122 [105,106], have been the object of previous theoretical studies.

The simulations, performed on the C-amidated and the N-acetylated peptide, were started from an ideal α-helix, a β-hairpin modelled on the basis of X-ray study [107] or from a completely extended conformations. The peptides were embedded in four different solvent models: water, tetrafluoroethanol (TFE), dimethysulfoxide (DMSO) and hexane. Moreover, as the HIS111 residue could be involved in pH-induced conformational changes of the prion protein in the pH 4–7 range [108,109], this residue was considered both singularly and doubly protonated to mimic neutral and acid conditions in water and TFE.

The simulations suggest that isolated PrP106-126 is mostly unstructured in solution [110]. The peptide can adopt a metastable helical structure in more apolar solvents, such as hexane (14%) and TFE (28%). In aqueous solution a mixture of β-sheet (18%) and helical structures (11%) were observed. In DMSO solution the peptide is effectively random coil. The range of helical structure in different environments is in rough agreement with that inferred from CD spectra. However, the highest probability of β-sheet formation in the simulations was observed in water, which was not supported experimentally.

In aqueous solution a persistent element of β-sheet between residues 108–112 and either residues 115–121 or 121–126 was observed. At neutral pH, an α–β transition was observed in the simulations. When His111 was protonated, mimicking acid conditions, no α–β transitions were observed and the probability of β-sheet formation decreased, in contrast with CD data obtained at pH 5 in buffer solution. Also the α-helical conformation in the region Ala113-Ala120 deduced by NMR experiment at pH 3.5 was not reproduced in the simulations. The $^{3}J_{HNH\alpha}$ coupling constants were also estimated and compared with experimental values for PrP106-126 in water and DMSO. Except for the region around residues 113–119, the agreement with the experimental data is reasonable in water. Experimentally, the ^{3}J coupling values in the 113–119 region are low (less than 6.0 Hz), indicating an α-helical conformation. In the simulations, α-helical elements were mainly observed in the hydrophilic regions. It is interesting to note that the J coupling constants calculated for residues 118–126 in water showed similar values despite whether His111 was singly (HIS) or doubly (HISH) protonated. This suggests that the protonation state of His111 does not significantly affect the conformational behaviour of the hydrophobic tail.

Thus, these MD simulations on PrP106-126 clearly indicate that the peptide shows conformational polymorphism in solution, giving support to the hypothesis of a possible role of this fragment in the structural transition of PrP^{C} to PrP^{Sc} in response to changes in the local environmental conditions.

The growing evidence of the existence, for each peptide, of environmental conditions (solvent, pH, temperature and concentration) that promote aggregation and fibril formation, independently of the amino acidic composition, feed an open debate about the nature of the interactions mainly driving such phenomena: interactions between backbones or between side chains? [111]

This evolving panorama stimulated us to extend our investigation to other sequences related to the 106–126 residues, with the aim of deepening the side chain contribution to secondary structure flexibility. At present we are working on a comparative molecular dynamics study on the conformational behaviour of the progressively growing sequences 113–120, 106–126 and 82–146 of the human prion protein in water, TFE and hexane.

For all three peptides, simulations starting with different initial velocities were conducted on α-helix or a completely extended conformation. For the 82–146 peptide we also started with the left handed β-helix, in line with the monomeric model proposed by Govaerts et al. [112]. This model was also used to build up a protofibril model [113] on the basis of electron microscopy data [90,91], of the truncated PrP^{Sc} (PrP27-30), and of the 106 amino acid fibrillogenic peptide termed miniprion ($PrP^{Sc}106$).

The palindrome sequence 113–120 (AGAAAAGA) falls within the hydrophobic tail of the 106–126 segment and is highly conserved in all the species for which the prion protein structure has been resolved. Furthermore, from molecular modelling studies there has emerged the possibility of its involvement in PrP^{C} to PrP^{Sc} transition. The 82–146 sequence [114] corresponds to the smallest amyolid unit purified from GSS brains and contains the segment 90–145, highly involved in fibrillogenic ability of the prion protein. In fact, residue 90 is the N-termination of the PrP27-30 peptide, which forms by limited proteolysis of the prion protein, and could be seen as the starter of the infective domain. Besides, the peptide of 106 amino acids termed PrP106 or miniprion, formed by residues 89–140 followed by residues 177–231, has been shown able to induce prion diseases in transgenic mice. It has also been observed that the deletion of the 90–145 residues prevents PrP^{Sc} formation. Finally, within this sequence there occurs a large part of the secondary structure changes implied in the fibrillogenic activity of the prion protein [115,116].

The conformational plasticity of these progressively growing peptides is monitored in both absolute and relative terms (i.e. the peptide as a whole and as part of a greater peptide) with the aim of discriminating between aminoacid specific effects and a sequence of cooperative effects on secondary structure flexibility in different environmental conditions. Preliminary results concerning the analysis of the simulations carried out on 113–120 and 106–126 peptides evidenced a reduced relative mobility of Gly114 and Gly119 in both the peptides, and similar behaviour appears associated with the two terminal glycines in the 106–126 sequence. From secondary structure data and principal components analysis, it seems that Gly113 and Gly119 act as a sort of structural pivot, as also emerged from experiments [117]. The 113–120 sequence seems to present specific flexibility, preserved also in the larger 106–126 segment, in accordance with the experimental observation that the palindrome sequence modulates not only the neurotoxic activity of the 106–126 peptide but also its secondary structure properties [118]. Definitive results, complete with an analysis of the simulations of the 82–146 peptide, will soon be presented.

The relation between the toxicity and chemical-physical properties of amyloid peptides pertains not only to prion diseases but also to other well known diseases such as Alzheimer's and Parkinson's disease that are characterized by the formation of amyloid aggregates as a consequence of conformational changes from the native form. Although there is some evidence concerning the toxicity of mature fibrils in these "conformational diseases", there are also studies that propose pre-fibrillar aggregates as primary toxic species. With the aim of contributing to a better understanding of the aggregation mechanism and the nature of such pre-fibrillar aggregates we are setting up, and testing, a procedure based on the use of Replica Exchange Molecular Dynamics simulations to overcome free energy traps on the multi peptide systems of the prion peptides 113–120, 106–126 and 82–146.

3.4 Quantum mechanical methods in bioinorganic chemistry: the activation of small molecules in hydrogenase and peroxidase enzyme pathways

The activation of small molecules (such as H_2, O_2, N_2, H_2O_2, NO, CO and CO_2) mediated by metal ions plays a central role in living organisms [119]. In addition, the elucidation of structure-function relationships in enzymes involved in small molecule processing is relevant not only to better understand the molecular basis of life processes, but also because it can drive the design of new and more efficient catalysts to be used in technological applications.

Among enzymes known to be involved in small molecule activation, we have focussed our attention on two families in recent years: hydrogenases and vanadium-peroxidases, which have attracted considerable attention due to their peculiar active-site features and their relevance in, respectively, H_2 production and oxidation chemistry.

Hydrogenases are enzymes that catalyze the reversible oxidation of dihydrogen, and invariantly contain transition metal ions. Two classes of hydrogenases have been extensively characterized. NiFe-hydrogenases are characterized by an active site where a nickel ion is coordinated by four cysteines, two of which are coordinated also to an iron ion. Biologically unusual ligands such as CO and CN complete the iron coordination environment. The active site of Fe-hydrogenases is formed by an unusual Fe_6S_6 cluster which is formed by a regular Fe_4S_4 cluster bridged by a cysteine residue to a binuclear subcluster where dihydrogen activation should take place [120,121].

Even though extensive experimental efforts aimed at disclosing the structural, electronic and reactivity characteristics of these enzymes have revealed many crucial aspects of hydrogenase chemistry [122] the elucidation of several key issues has benefited tremendously from theoretical studies. The reader is referred to a recent review for a full coverage of the topic [123].

Vanadium haloperoxidases (VHPOs) are enzymes that catalyze the oxidation of halide ions by hydrogen peroxide to the corresponding hypohalous acids. In addition, several peroxo-vanadium complexes have been shown to be good functional models of VHPO [124,125] and are capable of performing a variety of oxidation/oxygen transfer reactions. Some insight into the structural features of the inorganic cofactor and its environment has been obtained by X-ray diffraction of the VClPO from the pathogenic fungus *Curvularia Inaequalis* [126,127]. In the native state, the vanadium ion is characterized by trigonal bipyramidal geometry, where three oxygen atoms belong to the equatorial plane and one oxygen occupies an axial position. The other apical ligand is His496, which links the metal ion to the protein, whereas Lys353, Arg360, His404 and Arg490 are involved in hydrogen bonds with the oxygen atoms of the cofactor. In the peroxo derivative of the enzyme, the cofactor is characterized by strongly distorted trigonal bipyramidal geometry, with two oxo type

oxygen atoms and one peroxo atom in the equatorial plane, while His496 and the other peroxo atom occupy axial positions.

A thorough coverage of DFT implementation, strength and limitations was reviewed recently [128–134]. The hybrid three parameter B3LYP and the pure BP86 functionals have been the most widely used functionals to investigate hydrogenase and haloperoxidase models. In fact, both B3LYP and BP86 reproduced experimental geometries with high accuracy, and also reaction energies have been found in good agreement with experimental data, though in some cases (see below) the computed values were dramatically affected by the adopted functional. More specifically, in our recent studies, geometry optimizations and transition state searches have been carried out using the pure functional BP86 [135,136].

The first investigation in the field of hydrogenases carried out by our group was aimed at addressing the structural and electronic properties of the NiFe-hydrogenase active site in the different intermediate species formed in the catalytic cycle. In fact, several experimental approaches used to characterize paramagnetic intermediates could not be exploited for the characterization of diamagnetic forms, and, consequently, the structural and electronic properties of the Ni(II) intermediate species were even more elusive and controversial. We have shown that the μ-H species has the proper stereochemical features to correspond to the so-called Ni-C and Ni-R states [137,138]. Later, these conclusions were supported by other theoretical studies and also by experimental results [139]. According to our proposal, the heterolytic cleavage of dihydrogen takes place by a nucleophilic addition mechanism involving one of the cysteine ligands coordinated to the Ni centre.

Another controversial issue of NiFe-hydrogenase chemistry is related to the ground state of the Ni(II) ion in the Ni-SI and Ni-R intermediate species. In fact, spectroscopic data have, until recently, been considered compatible with a diamagnetic ground state for Ni(II). However, recently reported nickel L-edge soft X-ray data have suggested the involvement of high spin Ni(II) forms in the catalytic cycle, and this observation has stimulated more theoretical studies carried out within DFT. Moreover, the observation that B3LYP predicts the wrong multiplicity of the ground state for some transition metal complexes [140] has prompted efforts to test, and possibly tune, functionals for their use in the investigation of these issues [141–143]. In this context, we recently investigated the structural and electronic properties of high- and low-spin [Ni(II)S4] complexes, as well as Ni(II)Fe(II) models of the NiFe-hydrogenase active site, concluding that BP86 is well suited to describe the structural features of this class of compounds, whereas the prediction of the relative stability of low- and high-spin states is still problematic [144].

As for Fe-hydrogenases, with the aim of unravelling some of the key properties of the Fe-hydrogenase active site, we initially investigated the H_2, H^+ and H_2O adducts of $[(\mu\text{-PDT})Fe_2(CO)_3(CN)_2(CH_3S)]^-$, which is a model of the binuclear site found in the Fe-hydrogenase active site (Fig. 3) [145,146]. These studies contributed to the characterization of the reactivity properties of the binuclear cluster, leading to the conclusion that m-H species can be formed easily upon interaction between H^+ and H_2 with the metal cluster. Very recently, the investigation of intermediate species formed in the catalytic cycle was extended to models more closely resembling the enzyme active site. In particular, Density functional theory was used to compare reaction pathways for H_2 formation and H+ reduction catalyzed by models of the binuclear cluster found in the active site of [Fe] hydrogenases. Terminal H^+ binding to an FeI-FeI form, followed by monoelectron reduction and protonation of the di(thiomethyl)amine ligand, can conveniently lead to H_2 formation and release, suggesting that this mechanism could be operative within the enzyme active site. However, a pathway that implies the initial formation of FeII–FeII–H species and release of H_2 from the FeII–FeI form is characterized by only slightly less favoured energy

profiles. In both cases, H_2 formation becomes less favored when taking into account the competition between CN and amine groups for H^+ binding, an observation that can be relevant for the design of novel synthetic catalysts. H_2 cleavage can take place on FeII-FeII redox species, in agreement with previous proposals [147] and, in complexes characterized by terminal CO groups, does not need the involvement of an external base. The step in H_2 oxidation characterized by larger energy barriers corresponds to the second H^+ extraction from the cluster, both considering FeII–FeII and FeII–FeIII species. A comparison of the different reaction pathways reveals that H_2 formation could involve only FeI–FeI, FeII–FeI, and FeII–FeII species, whereas FeIII–FeII species might be relevant in H_2 cleavage.

Another puzzling aspect related to the structure of the Fe-hydrogenase active site is the X-ray structure of the fully reduced binuclear cluster. In fact, biomimetic models are generally characterized by structures featuring terminal CO groups and compatible with an Fe–Fe bond, whereas in the enzyme a CO group bridges the metal centre and, consequently, a vacant coordination site where H_2 is thought to bind appears on the iron centre distal to the Fe_4S_4 cluster. To contribute to addressing this issue we studied the factors affecting the structural and electronic properties in a series of complexes characterized by the general structure $[Fe_2(\mu\text{-PDT})(CO)_3(CN)_2L]$, where the ligand L was systematically replaced by groups with different s-donor and p-acceptor character. The results led to the conclusion that the substitution in the parent complex $[Fe_2(\mu\text{-PDT})(CO)_6]$ of CO ligands with two CN and an electron donor L ligand is sufficient to modify the structure of the bimetallic cluster from a pseudo-symmetrical edge-bridged square pyramid to a μ-CO species characterized by a vacant coordination site on the distal Fe ion. The influence of the electronic properties of L on the chemistry of the bimetallic cluster suggested also that the oxidation state of the [Fe_4S_4 cluster] could affect the structural and electronic properties of the binuclear cluster, possibly modulating the relative stability of μ-CO and CO-unbridged species [148].

In addition, we also recently reported the DFT characterization of Fe6S6 models of the whole cluster found in the active site of Fe-hydrogenases [149], as well as more extended models of the active site [150]. Along these lines, QM/MM investigations of the protein systems are currently underway in collaboration with Professor Ryde (Lund University, Sweden).

DFT was also used by our group to address problems related to the reactivity of synthetic biomimetic hydrogenase models. As an example, the systematic experimental and computational study of diferrous dicyano dithiolates showed that the oxidation of $[Fe_2(S_2C_2H_4)\ (CN)_2(CO)_4]^{2-}$ in the presence of cyanide and tertiary phosphines and of $Fe_2(S_2C_2H_4)(CO)_4(PMe_3)_2$ in the presence of cyanide affords a series of diferrous cyanide derivatives that bear a stoichiometric, structural, and electronic relationship oxidized state of Fe-only hydrogenases. DFT calculations revealed that the most stable isomers of $Fe_2(S_2C_2H_4)(\mu\text{-CO})(CN)_2(PMe_3)_2\ (CO)_2$ have cyanide trans to μ-CO, an relevant observation to understand the properties of the CO inhibited form of the enzyme [151].

Density functional theory was also used to investigate the reaction between H_2 and [Ni(NHPnPr$_3$)(‘S3’)] or [Pd(NHPnPr$_3$)(‘S3’)], where ‘S3’ = bis(2-sulfanyl-phenyl) sulfide(2−), which are among the few synthetic complexes featuring a metal coordination environment similar to that observed in the [NiFe] hydrogenase active site and capable of catalyzing H_2 heterolytic cleavage. The results allowed us to unravel the reaction mechanism, which is consistent with an oxidative addition-hydrogen migration pathway for [Ni(NHPnPr$_3$)(‘S3’)], whereas metathesis is also possible with [Pd(NHPnPr$_3$) (‘S3’)]. Unexpectedly, H_2 binding and activation implies structural reorganization of the metal coordination environment. It turns out that the structural rearrangement in [Ni(NHPnPr$_3$) (‘S3’)] and [Pd(NHPnPr$_3$)(‘S3’)] can take place due to the peculiar structural features of the

Ni and Pd ligands, explaining the remarkable catalytic properties. However, the structural reorganization is the most unfavourable step along the H_2 cleavage pathway ($\Delta G > 100$ kJ mol^{-1}), an observation that is relevant for the design and synthesis of novel biomimetic catalysts [152].

In another conceptually similar study, we used the density functional theory (DFT) to dissect the overall CN/CO substitution pathway on the {2Fe3S} complex $[Fe_2(CO)_5\{MeSCH_2C(Me)(CH_2S)_2\}]$, in terms of the energetics and structures of transition states, intermediates and products. We show that the formation of bridging CO transition states is explicitly involved in the intimate mechanism of dicyanation. The enhanced rate of monocyanation of {2Fe3S} over the {2Fe2S} species $[Fe_2(CO)_6\{CH_2(CH_2S)_2\}]$ was found to rest with the ability of the thioether ligand to both stabilize a m-CO transition state and act as a good leaving group. In contrast, the second cyanation step of the {2Fe3S} species is kinetically slower than for the {2Fe2S} monocyanide because one of the Fe atoms is deactivated by the coordination of the electron-donating thioether group. It also turned out that the intermediate species formed in the second cyanation step of {2Fe3S} species is a μ-CO species, confirming the structural assignment made on the basis of FT-IR data [153]. In a wider context, the study provides some insight into the reactivity of dinuclear systems in which neighbouring group on-off coordination can play a role in substitution pathways [154].

To shed light on structure–function relationships in the vanadium haloperoxidase enzyme family, we recently used DFT to investigate the structural and electronic properties of complexes related to the resting form of the active site of vanadium haloperoxidase, as a function of environment and the protonation state. The results highlight the influence of environment and the protonation state on the structure and stability of the metal cofactor. The study showed that, in the trigonal bipyramidal active site where one axial position is occupied by a key histidine, the trans position cannot contain a terminal oxo group. Moreover, protonation of at least one equatorial oxo ligand appears necessary to stabilize the metal cofactor. The study also indicated that, while at rest within the protein, the vanadate unit is most likely an anion with an axial hydroxide and an equatorial plane containing two oxos and a hydroxide. For the neutral, protonated state of the vanadate unit, two isomers were characterized. The first structure features an axial water with two oxo and one hydroxo group in the equatorial plane. The second structure contains an axial hydroxo group and an equatorial plane composed of one oxo and two hydroxo oxygen atoms. These two species are not significantly different in energy, indicating that either form may be important during the catalytic cycle [155].

In a following study, DFT was used to investigate structural, electronic and reactivity properties of complexes related to the peroxo forms of vanadium haloperoxidases. In particular, the reactivity of the cofactor as a function of protonation state and environment, which are two factors thought to be crucial in modulating the activity of the enzyme, was examined. The results highlighted the role of protonation in the activation of the peroxo-vanadium complexes, and showed that the oxo-transfer step involves the unprotonated axial peroxo oxygen atom which is easily accessible to substrates in the peroxo form of the enzyme. The role of Lys353, which in the X-ray structure of the peroxide-bound form of vanadium chloroperoxidase is hydrogen bonded to the equatorial oxygen atom of the peroxo group, was also explored, leading to the conclusion that Lys353 can play a role similar to a H^+ in the activation of the peroxo form of the cofactor (Fig. 4) [156].

We are presently extending our DFT investigation of the catalytic properties of VHPO with the aim of characterizing other relevant intermediate species formed in the catalytic cycle. To this end we are planning to study more extended models, using QM/MM approaches, and

compare experimental spectroscopic data with the corresponding computational results obtained within the time-dependent DFT theoretical framework.

3.5 Developing and applying suitable strategies for variable reduction in 3D-QSAR investigations: the defibrillogenic activity of tetracyclines

A number of methodologies for Three Dimensional Quantitative Structure Activity Relationships (3D-QSAR) that offer the possibility of dealing with ligand–receptor interactions in three dimensions have been developed [10–17]. These methodologies can be classified as global or local, depending on the descriptors selected to outline the characteristics of the 3D distribution of molecular interaction fields.

Global approaches are based on a few descriptors that summarize the field characteristics. In this framework, we developed a new methodology, based on the Grid Weighted Holistic Invariant Molecular (G-WHIM) descriptors, that summarize all the information of the whole field distribution in terms of a small number of dimension and shape indexes [157]. A drawback of global approaches is that the relationships between biological activity and the field values in the different regions around the molecule cannot be shown, thus it is often difficult to give a physical meaning to the model.

On the contrary, this information is contained in the local models where the QSAR descriptors are the values assumed by the interaction field in the 3D grid embedding the molecule. Among the local approaches, the CoMFA method [158] is certainly the most popular. Molecular alignment is probably the most crucial problem of local methods in 3D-QSAR analyses, as all these methods require an alignment criterion before developing the quantitative model. Poor alignment can result in an inadequate statistical model. An underlying assumption in QSAR analyses is that all molecules in the data set showing high activity bind to their receptor in a similar way. If the molecules present similar molecular skeletons or similar binding groups, molecular alignment can be performed by skeleton or binding group superimposition i.e. the pharmacophore using the most active compound in the series as a template. Alternatively, the alignment can be performed on the basis of similarities in 3D interaction fields.

In addition to this problem, the local methods must consider thousands of descriptors, i.e. the property values calculated in the 3D grid where the molecule is embedded. However, there is a conflict between using many grid points to produce a more accurate description of the interaction field and the notion of keeping the number of variables low to reduce the noise in the model. Well-suited strategies have been proposed [23,24] to select only those descriptors likely to yield regression models with predictive ability. However, the high number of independent variables makes it difficult to extract chemical information relevant to the quantitative structure-activity problem.

To overcome this problem, we proposed a new local variable reduction technique (named Statistical Extrema for Structure Activity Model Evaluation, SESAME) [25], based on the selection of grid points where suitable statistical indexes of the interaction potential distributions assume a local maximum (or minimum) value. The standard deviation (SD) of the interaction potential values of all the molecules in the grid points, as well as any correlation index between the potential values and the biological activities can be considered. Among the different correlation indexes, the Pearson coefficient, the Spearman rank correlation index and the entropy of Shannon can be used. The index used depends on the activity data: the Pearson coefficient can be used when quantitative activity data are available; the Spearman index, when only activity ranks are available; the entropy of Shannon, when only classes of activity are defined.

The methodology was initially applied to the steroid data set, often used as a benchmark to compare new 3D QSAR methodologies [17].

Briefly the methodology can be outlined as follows. For a set of properly aligned molecules the Molecular Electrostatic Potential (MEP) and the non-bonded (van der Waals) interaction energy values are calculated in a 3D grid of points embedding all the molecules. Points within the surface, obtained as the envelopment of the van der Waals surfaces of all the molecules, are neglected; at each point outside the surface, proper statistical indexes are calculated, i.e. the standard deviation of the interaction potentials and/or the correlation between the vector of the interaction potential values generated by each molecule and the vector of the biological activities.

To avoid the selection of physically meaningless points, and to retain only relevant descriptors for QSAR modelling, selection is performed by retaining only those grid points that satisfy the following conditions: (a) the standard deviation of the interaction potential values must be greater than a predefined threshold; (b) the statistical index value must be greater or equal to a pre-defined threshold; (c) the statistical index must be a local maximum or minimum with respect to the first and second nearest points. Thus, the reduction procedure selects only those grid points likely to yield regression models with predictive ability as independent variables for the QSAR models.

The best least squares regression models are searched for by using the Genetic Algorithm-Variable Subset Selection (GA-VSS) approach [54,55]. This method can individualize the best correlation models through a selection of the independent variables truly correlated with the response variable. Selection is performed by the leave-one-out cross validated regression coefficient, Q^2_{LOO}. Finally, a fully cross-validated regression coefficient, $Q^2_{20\%}$, with 20% cancellation groups, is calculated for the most effective models.

The variables that appear in the final model allow the identification of the regions around the molecules that are most relevant for the binding process.

Here we present the preliminary results of a 3D-QSAR study carried out to identify the stereoelectronic features required for the observed defibrillogenic activity toward aggregates of the PrP106-126 peptide for tetracycline itself and 14 tetracycline derivatives.

Recently, different classes of compounds [159] were identified as being able to antagonize prion propagation in cellular and/or animal models of the Transmissible spongiform encephalopathies (TSEs), or prion disease. Among these, tetracycline and its analogues were found to affect prion infectivity [160] and present in vitro anti-amyloidogenic activity [161]. Tetracycline was found, by fluorescence and NMR spectroscopy, to bind to PrP aggregates and to oligomeric forms of synthetic PrP peptides [162]. Moreover, the ability of tetracycline to prevent aggregation and acquisition of protease resistance of PrP peptides, and to disrupt PrP peptides aggregates [162], was observed.

Tetracyclines are a group of structurally-related antibiotics with well characterized pharmacological and pharmacokinetic properties [163]. These molecules have five asymmetric centres on a common hydronaphthacene moiety containing four fused rings [164,165]. The main features required for antibacterial activity are well established [163] and, according to these, the clinically used tetracyclines present various substitutions at the 5, 6 and 7 positions.

The chemistry of tetracyclines in solution is quite complicated due to their ability to adopt different protonation states, tautomeric forms, and conformations depending on the pH, the presence of metal cations, and the solvent [164–167].

At present, the defibrillogenic mechanism of action of tetracycline derivatives is unknown. A better understanding of the stereoelectronic features of tetracyclines involved in their anti-prion activity could help characterize their action mechanism, and may lead to the design and synthesis of new analogues with better anti-amyloidogenic, pharmacological and financial profiles and devoid of antibiotic effects.

The conformational space of the considered tetracy-clines was previously investigated [167] in aqueous solution by the Monte Carlo Multiple Minimum method [168,169], using the MM2* force field in combination with the generalized Born-solvent accessible surface area continuum solvation model [170,171], as implemented in the MacroModel 7.2 release [172] within the Maestro molecular modelling interface [173]. All the compounds were considered in their neutral protonation state and in the zwitterionic form, in which the hydroxyl group at C_3 is deprotonated while the adjacent 4-Nme_2 group is protonated. An analysis of the conformational flexibility of the tetracycline derivatives did not provide a geometrical pharmacophore model able to justify the differences observed in the defibrillogenic activity; for this reason, we performed a 3D-QSAR analysis of the tetracycline defibrillogenic activity on the PrP106-126 aggregates.

The variable reduction step, achieved by *SESAME* on 200,000 3D grid points, provided 173 points where the Pearson correlation coefficient assumed local maxima or minima: the interaction potential values at these points were used as descriptors in the search for the best QSAR models.

The population of best models obtained by GA-VSS was then analysed, searching for models with high predictive power, physical meaning, and stability to the increase of the number of variables contained in the models. One set of models presents the desired properties and exhibits good statistical quality, both in fitting and prediction, in the calculated values of defibrillogenic activity (the Q^2_{LOO} values range between 0.970 and 0.927). Model quality increases with increasing numbers of variables, and the models are stable to the increase in the number of variables, i.e. the descriptors present in the best 4-variable model are also included in the best 5-variable model, and so on. Moreover, these models include points with maxima or minima values of R_{VDW} and R_{MEP}, highlighting the relevance of both steric and electrostatic terms in tuning the defibrillogenic activity of the molecules.

Inspection of the descriptors entering the models allowed us to highlight the influence several structural and electronic parameters have on defibrillogenic activity (Fig. 5): (a) the electronic effects of substitution at the aromatic ring ($R_{\mathrm{MEP}}1$); (b) the electronic effect of the hydroxyl group at position C_5($R_{\mathrm{MEP}}2$); the steric effect of the hydroxyl group at position C_6($R_{\mathrm{VDW}}3$, 5, 6); (c) the steric hindrance arising from the bulky dimethylammonium group when the stereochemical configuration at C_4 is *epi* ($R_{\mathrm{VDW}}4$).

In conclusion, these models are able to correlate the electronic and steric differences of the 15 considered tetracyclines with defibrillogenic activity. Thus, these models could be used to predict new tetracycline derivatives with enhanced defibrillogenic activity and be used for future drug development.

4 Conclusions

Several biological processes have been faced at the molecular level using different computational approaches, ranging from Molecular Dynamics simulations, Quantum Mechanical methods and 3D Quantitative Structure–Activity Relationships. The findings highlight the capability of the adopted methodologies to bring important improvements to the elucidation of the mechanisms underlying the different biological processes.

The information obtainable from the different methods should appear better integrated in future work, particularly the combined use of Molecular Dynamics and homology modelling methods which appears very promising towards finding structure–function relationships in specific protein families.

Quantum mechanical calculations in the field of Bio-inorganic Chemistry will particularly address the search for suitable QM/MM protocols that will also include the long range effects caused by the protein environment. Indeed, studies are presently in progress, the aim being to obtain an accurate description of the enzyme reaction path.

Finally, structural prediction methods could be used to provide molecular models which would validate QSAR results when protein structure is lacking.

Acknowledgments

We gratefully acknowledge financial support from the Italian Ministry of University and Research (PRIN 2003), CINECA (grants no. 339068763168 and no. 460130052612) and the National Institutes of Environmental Health Sciences (grant ES07685).

References

1. Special issue on Molecular Dynamics Simulations of Bio-molecules. Acc Chem Res 2002;35(6)
2. Rousseau F, Schymkowitz J. Curr Opin Struct Biol 2005;15:23. [PubMed: 15718129]
3. Tama F. Protein Peptide Lett 2003;10:119.
4. Bhattacharyya K. Acc Chem Res 2003;36:95. [PubMed: 12589694]
5. Noodleman L, Lovell T, Han WG, Li J, Himo F. Chem Rev 2004;104:459. [PubMed: 14871132]
6. Himo F, Siegbahn PEM. Chem Rev 2003;103:2421. [PubMed: 12797836]
7. Siegbahn PEM. Blomberg MRA. Chem Rev 2000;100:421. [PubMed: 11749242]
8. Dreuw A, Head-Gordon M. Chem Rev 2005;105:4009. [PubMed: 16277369]
9. Carter EA, Rossky PJ. Acc Chem Res 2006;39:71.
10. Hasegawa K, Arakawa M, Funatsu K. Curr Comput-Aided Drug Des 2005;1:129.
11. Walters WP, Goldman BB. Curr Opin Drug Dis Dev 2005;8:329.
12. Murray JS, Politzer P. Methods Princ Med Chem 2003;17:233.
13. Akamatsu M. Curr Top Med Chem 2002;2:1381. [PubMed: 12470286]
14. Hansch C, Hoekman D, Leo A, Weininger D, Selassie CD. Chem Rev 2002;102:783. [PubMed: 11890757]
15. Patrick, GL. An introduction to medicinal chemistry. 3. Oxford University Press; Oxford: 2005.
16. Oprea, TI.; Waller, CL. Rev Comput Chem. Lipkowitz, KB.; Boyd, DB., editors. Vol. 11. VHC Publishers; New York: 1997. p. 127and references therein
17. Coats, EA. Drug Discovery Des. Vol. 12. Kluwer Academic Publishers; UK: 1998. Perspect; p. 199
18. Amadei A, Linssen AB, Berendsen HJC. Proteins 1993;17:412. [PubMed: 8108382]
19. Lin, H.; Truhlar, DG. Theoretical Chemistry Accounts. 2006. online first: 1, and references therein
20. Friesner RA, Baik M-H, Gherman BF, Guallar V, Wirstam M, Murphy RB, Lippard SJ. Coordination Chem Rev 2003;238–239:267.

21. Ridder L, Mulholland AJ. Curr Top Med Chem 2003;3:1241. [PubMed: 12769703]
22. Mulholland AJ. Theor Comput Chem 2001;9:597.
23. Baroni M, Costantino G, Cruciani G, Riganelli D, Valigi R, Clementi S. Quant Struct Act Relat 1993;12:9.
24. Lindgren F, Geladi P, Rännar S, Wold S. J Chemom 1994;8:349.
25. Cosentino U, Moro G, Bonalumi D, Bonati L, Lasagni M, Todeschini R, Pitea D. Chemom Intell Lab Syst 2000;52:183.
26. Holm L, Sander C. J Mol Biol 1993;233:123. [PubMed: 8377180]
27. Thompson JD, Higgins DG, Gibson TJ. Nucleic Acids Res 1994;22:4673. [PubMed: 7984417]
28. Sali A, Blundell TL. J Mol Biol 1993;234:779. [PubMed: 8254673]
29. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. J Appl Cryst 1993;26:283.
30. Sippl MJ. Proteins 1993;17:355. [PubMed: 8108378]
31. DeLano, WL. The PyMOL molecular graphics system. DeLano Scientific; San Carlos, CA, USA: 2002.
32. Kelley LA, MacCallum RM, Sternberg MJE. J Mol Biol 2000;299:499. [PubMed: 10860755]
33. Jones DT. J Mol Biol 1999;287:797. [PubMed: 10191147]
34. Alexandrov, N.; Nussinov, R.; Zimmer, R. Pac Symp Biocomput. Hunter, L.; Klein, TE., editors. World Scientific Publishing Co; Singapore: 1996. p. 53
35. Shi J, Blundell TL, Mizuguchi K. J Biol Mol 2001;301:243.
36. Rost, B. The third international conference on intelligent system for molecular biology (ISMB). Rawlings, C.; Clark, D.; Altman, R.; Hunter, L.; Lengauer, T.; Wodak, S., editors. AAAI Press; Cambridge, UK, Menlo Park, CA: 1995. p. 314
37. Berendsen HJC, van der Spoel D, van Drunen R. Comput Phys Comm 1995;91:43.
38. Lindahl E, Hess B, van der Spoel D. J Mol Mod 2001;7:306.
39. Berendsen, HJC.; Postma, JPM.; van Gusteren, WF.; Hermans, J. Interaction models for water in relation to protein hydratation. Reidel; Dordrecht: 1981. p. 331
40. Berendsen HJC, Postma JPM, van Gusteren WF, Di Nola A, Haak JR. J Chem Phys 1984;81:3684.
41. Darden TA, York DM, Pedersen LG. J Chem Phys 1993;98:10089.
42. Hess B, Bekker H, Berendsen HJC, Fraahje JGEM. J Comp Chem 1997;18:1463.
43. Ahlrichs R, Bar M, Haser M, Horn H, Kolmel C. Chem Phys Lett 1989;162:165.
44. Schafer A, Huber C, Ahlrichs R. J Chem Phys 1994;100:5829.
45. Jensen, F. Introduction to computational chemistry. Wiley; Baffins Lane, Chichester, England: 2000.
46. Klamt A. J Phys Chem 1995;99:2224.
47. Klamt A. J Phys Chem 1996;100:3349.
48. Himo F, Noodleman L, Blomberg MRA, Siegbahn PEM. J Phys Chem A 2002;106:8757.
49. Klamnt A. J Phys Chem 1995;99:2224.
50. Klamnt A. J Phys Chem 1996;100:3349.
51. Warshel A, Naray-Szabo G, Sussman F, Hwang J-K. Biochemistry 1989;28:3629. [PubMed: 2665806]
52. Bottoni A, Lanza CZ, Miscione GP, Spinelli D. J Am Chem Soc 2004;126:1542. [PubMed: 14759213]
53. Clark M, Cramer RD III, Van Opdenbosch N. J Comp Chem 1989;10:982.
54. Goldberg, DE. Genetic algorithms in search, optimization and machine learning. Addison-Wesley; Massachusetts: 1989.
55. Leardi R, Boggia R, Terrile M. J Chemom 1992;6:267.
56. Safe S. Crit Rev Toxicol 1990;21:51. [PubMed: 2124811]
57. Schmidt JV, Bradfield CA. Annu Rev Cell Dev Biol 1996;12:55. [PubMed: 8970722]
58. Procopio M, Lahm A, Tramontano A, Bonati L, Pitea D. Eur J Biochem 2002;269:13. [PubMed: 11784293]

59. Denison MS, Pandini A, Nagy SR, Baldwin EP, Bonati L. Chem Biol Interact 2002;141:3. [PubMed: 12213382]
60. Gong W, Hao B, Mansy SS, Gonzalez G, Gilles-Gonzalez MA, Chan MK. Proc Natl Acad Sci USA 1998;95:15177. [PubMed: 9860942]
61. Murray IA, Reen RK, Leathery N, Ramadoss P, Bonati L, Gonzalez FJ, Peters JM, Perdew GH. Archives Biochem Biophys 2005;442:59.
62. Erbel PJ, Card PB, Karakuzu O, Bruick RK, Gardner KH. Proc Natl Acad Sci USA 2003;100:15504. [PubMed: 14668441]
63. Card PB, Erbel PJ, Gardner KH. J Mol Biol 2005;353:664. [PubMed: 16181639]
64. Denison MS, Soshilov A, Song Y, Pandini A, Bonati L. Organohalogen Compounds 2005:114.
65. Pandini A, Denison MS, Song Y, Soshilov AA, Bonati L. Biochemistry. (in press).
66. Pandini A, Bonati L. Protein Eng Des Sel 2005;18:127. [PubMed: 15820977]
67. Hess B. Phys Rev E 2000;62:8438.
68. Hess B. Phys Rev E 2002;65:031910–1.
69. Hellingwerf KJ, Hendriks J, Gensch T. J Phys Chem 2003;107:1082.
70. Amezcua C, Harper S, Rutter J, Gardner K. Structure (Camb) 2002;10:1349. [PubMed: 12377121]
71. Crosson S, Moffat K. Plant Cell 2002;14:1067. [PubMed: 12034897]
72. Pandini A, Mauri G, Bonati L. (Submitted).
73. Hedstrom L. Chem Rev 2002;102:4501. [PubMed: 12475199]
74. Papaleo E, Fantucci P, De Gioia L. J Chem Theory Comput 2005;1:1286.
75. Sondermann H, Soisson SM, Boykevisch S, Yang S-S, Bar-Sagi D, Kuriyan J. Cell 2004;119:393. [PubMed: 15507210]
76. Greco C, Fantucci P, De Gioia L. BMC Bioinformatics 2005;6:S15. [PubMed: 16351741]
77. Sacco E, Fantinato S, Manzoni R, Metalli D, De Gioia L, Fantucci P, Alberghina L, Vanoni M. FEBS Lett 2005;579:6851. [PubMed: 16325815]
78. Mereghetti P, Papaleo E, Fantucci P, De Gioia L. (Submitted).
79. Prusiner SB. Science 1991;252:1515. [PubMed: 1675487]
80. Prusiner SB, Scott MR, De Armond SJ, Cohen FE. Cell 1998;93:337. [PubMed: 9590169]
81. De Armond SJ, Yang SL, Lee A, Bowler R, Tarabolous A, Groth D, Prusiner SB. Proc Natl Acad Sci USA 1993;90:6449. [PubMed: 8101989]
82. Williams A, Lucassen PJ, Ritchie D, Bruce M. Expt Neurol 1997;14:433.
83. Stefani M, Dobson CM. J Mol Med 2003;81:678. [PubMed: 12942175]
84. Caughey BW, Dong A, Bhat KS, Ernst D, Hayes SF, Caughey WS. Biochemistry 1991;30:7672. [PubMed: 1678278]
85. Pan KM, Baldwin M, Nguyen J, Gasset M, Serban A, Groth D, Mehlhorn I, Huang Z, Fletterick RJ, Cohen FE, Prusiner SB. Proc Natl Acad Sci USA 1993;90:10962. [PubMed: 7902575]
86. James TL, Liu H, Ulyanov NB, Farr-Jones S, Zhang H, Donne DG, Kaneko K, Groth D, Mehlhorn I, Prusiner SB, Cohen FE. Proc Natl Acad Sci USA 1997;94:10086. [PubMed: 9294167]
87. Donne DG, Viles JH, Groth D, Mehlhorn I, James TL, Cohen FE, Prusiner SB, Wright PE, Dyson HJ. Proc Natl Acad Sci USA 1997;94:13452. [PubMed: 9391046]
88. Zhan R, Liu H, Lührs T, Riek R, von Schroetter C, Garcia L, Billiter M, Calzolai L, Wider G, Wüthrich K. Proc Natl Acad Sci USA 2000;97:145. [PubMed: 10618385]
89. Wille H, Zhang GF, Baldwin MA, Cohen FE, Prusiner SB. J Mol Biol 1996;259:608. [PubMed: 8683568]
90. Wille H, Michelitsch MD, Guenebaut V, Supattopone S, Serban A, Cohen FE, Agard DA, Prusiner SB. Proc Natl Acad Sci USA 2001;98:11686. [PubMed: 11562491]
91. Wille H, Michelitsch MD, Guènebaut V, Supattopone S, Serban A, Cohen FE, Agard DA, Prusiner SB. Proc Natl Acad Sci USA 2002;99:3563. [PubMed: 11891310]
92. Peretz D, Williamson RA, Matsunaga Y, Serban H, Pinilla C, Bastidas RB, Rozenhsteyn R, James TL, Houghten RA, Cohen FE, et al. J Mol Biol 1997;273:614. [PubMed: 9356250]

93. Williamson RA, Peretz D, Pinilla C, Ball H, Bastidas RB, Rozenshteyn R, Houghten RA, Prusiner SB, Burton DR. J Virol 1998;72:9413. [PubMed: 9765500]
94. Forloni G, Angeretti N, Chiesa R, Monzani E, Salmona M, Bugiani O, Tagliavini F. Nature 1993;362:543. [PubMed: 8464494]
95. Brown DR, Schmidt B, Kretzschmer HA. Nature 1996;380:345. [PubMed: 8598929]
96. Forloni G, Del Bo R, Angeretti N, Chiesa R, Smiroldo S, Doni R, Ghibaudi E, Salmona M, Porro M, Verga L, Giaccone G, Bugiani O, Tagliavini F. Eur J Neurosci 1994;6:1415. [PubMed: 8000566]
97. Tagliavini F, Prelli F, Verga L, Giaccone G, Sarma R, Gorevic P, Ghetti B, Passerini F, Ghibaudi E, Forloni G, Salmona M, Bugiani O, Frangioni B. Proc Natl Acad Sci USA 1993;90:9678. [PubMed: 8105481]
98. Tagliavini F, Prelli F, Porro M, Rossi G, Giaccone G, Farlow MR, Dlouhy SR, Ghetti B, Bugiani O, Frangione B. Cell 1994;79:695. [PubMed: 7954833]
99. Selvaggini C, De Gioa L, Cantù L, Ghibaudi E, Diomede L, Passerini F, Forloni G, Bugiani O, Tagliavini F, Salmona M. Biochem Biophys Res Commun 1993;194:1380. [PubMed: 8102526]
100. De Gioia L, Selvaggini C, Ghibaudi E, Diomede L, Bugiani O, Forloni G, Tagliavini F, Salmona M. J Biol Chem 1994;269:7859. [PubMed: 7907586]
101. Salmona M, Malesani P, De Gioia L, Gorla S, Bruschi M, Molinari A DellaVedova R, Pedrotti B, Marrari MA, Awan T, Bugiani O, Forloni G, Tagliavini F. Biochem J 1999;342:207. [PubMed: 10432318]
102. Ragg E, Tagliavini F, Malesani P, Monticelli L, Bugiani O, Forloni G, Salmona M. Eur J Biochem 1999;266:1192. [PubMed: 10583417]
103. Kuwata K, Matumoto T, Cheng H, Nagayama K, James TL, Roder H. Proc Natl Acad Sci USA 2003;100:14790. [PubMed: 14657385]
104. Levy Y, Hanan E, Solomon B, Becker OM. Proteins 2001;45:382. [PubMed: 11746686]
105. Kazmirski SL, Alonso DOV, Cohen FE, Prusiner SB, Daggett V. Chem Biol 1995;2:305. [PubMed: 9383432]
106. Daidone I, Simona F, Roccatano D, Broglia RA, Tiana G, Colombo G, Di Nola A. Proteins 2004;57:198. [PubMed: 15326604]
107. Inouye H, Kirschner DA. J Struct Biol 1998;122:247. [PubMed: 9724626]
108. Calzolai L, Zahn R. J Biol Chem 2003;278:35592–35596. [PubMed: 12826672]
109. Langella E, Improta R, Crescenzi O, Barone V. Proteins 2006;64:167. [PubMed: 16639746]
110. Villa A, Mark AE, Saracino GAA, Cosentino U, Pitea D, Moro G, Salmona M. J Phys Chem B 2006;110:1423. [PubMed: 16471693]
111. Dobson CM. Nature 2003;426:884. [PubMed: 14685248]
112. Govaerts C, Wille H, Prusiner SB, Cohen FE. Proc Natl Acad Sci USA 2004;101:8342. [PubMed: 15155909]
113. DeMarco ML, Daggett V. Proc Natl Acad Sci USA 2004;101:2293. [PubMed: 14983003]
114. Gobbi M, Colombo L, Morbin M, Mazzoleni G, Accardo E, Vanoni M, Del Bavero E, Cantù L, Kirschner DA, Manzoni C, Beeg M, Ceci P, Ubezio P, Forloni G, Tagliavini F, Salmona M. J Biol Chem 2006;281:843. [PubMed: 16286452]
115. Laws DD, Bitter HML, Liu K, Ball HL, Kaneko K, Wille H, Cohen FE, Prusiner SB, Pines A, Wemmer DE. Proc Natl Acad Sci USA 2001;98:11686. [PubMed: 11562491]
116. Supattopone S, Bosque P, Muramoto T, Wille H, Aagaard C, Peretz D, Nguyen HOB, Heinrich C, Torchia M, Safar J, Cohen FE, DeArmond SJ, Prusiner SB, Scott M. Cell 1999;96:869. [PubMed: 10102274]
117. Florio T, Paludi D, Villa V, Rossi Principe D, Corsaro A, Millo E, Damonte G, D'Arrigo C, Russo C, Schettini G, Aceto A. J Neurochem 2003;85:62. [PubMed: 12641727]
118. Jobling MF, Stewart AR, White AR, McLean C, Friedhuber A, Maher F, Beyereuther K, Masters CL, Barrow CL, Collins SJ, Cappai R. J Neurochem 1999;73:1557. [PubMed: 10501201]
119. Macyk W, Franke A, Stochel G. Coord Chem Rev 2005;249:2437.
120. Armstrong FA. Curr Opin Chem Biol 2004;8:133. [PubMed: 15062773]

121. Evans DJ, Pickett CJ. Chem Soc Rev 2003;32:268. [PubMed: 14518180]
122. Volbeda A, Fontecilla-Camps JC. Coord Chem Rev 2005;249:1609.
123. Bruschi M, Zampella G, Fantucci P, De Gioia L. Coord Chem Rev 2005;249:1620.
124. Butler A, Clague MJ, Meister GE. Chem Rev 1994;94:625.
125. Plass W. Coord Chem Rev 2003;237:205.
126. Messerschmidt A, Wever R. Proc Natl Acad Sci USA 1996;93:392. [PubMed: 8552646]
127. Messerschmidt A, Prade L, Wever R. Biol Chem 1997;378:309. [PubMed: 9165086]
128. Ban F, Rankin KN, Gauld JW, Boyd RJ. Theoretical Chemistry Accounts 2002;108:1.
129. Friesner RA. Proc Natl Acad Sci USA 2005;102:6648. [PubMed: 15870212]
130. Niu S, Hall MB. Chem Rev 2000;100:353. [PubMed: 11749240]
131. Siegbahn PEM, Blomberg MRA. Chem Rev 2000;100:421. [PubMed: 11749242]
132. Lovell T, Himo F, Han W-G, Noodleman L. Coord Chem Rev 2003;238–239:211.
133. Friesner RA, Baik M-H, Gherman BF, Guallar V, Wirstam M, Murphy RB, Lippard SJ. Coord Chem Rev 2003;238–239:267.
134. Improta R, Barone V. J Comput Chem 2004;25:1333. [PubMed: 15185326]
135. Becke AD. Phys Rev A 1988;38:3098. [PubMed: 9900728]
136. Perdew JP. Phys Rev B 1986;33:8822.
137. De Gioia L, Fantucci P, Guigliarelli B, Bertrand P. Inorg Chem 1999;38:2658.
138. De Gioia L, Fantucci P, Guigliarelli B, Bertrand P. Int J Quantum Chem 1999;73:187.
139. Stein M, Lubitz W. Curr Opin Chem Biol 2002;6:243. [PubMed: 12039011]
140. Reiher M, Salomon O, Sellmann D, Hess BA. Chem Eur J 2001;7:5195.
141. Cavallo L, Jacobsen H. J Phys Chem A 2003;107:5466.
142. Paulsen H, Duelund L, Winkler H, Toftlund H, Trautwein AX. Inorg Chem 2001;40:2201. [PubMed: 11304167]
143. Poli R, Harvey JN. Chem Soc Rev 2003;32:1. [PubMed: 12596540]
144. Bruschi M, De Gioia L, Zampella G, Reiher M, Fantucci P, Stein M. J Biol Inorg Chem 2004;9:873. [PubMed: 15365900]
145. Bruschi M, Fantucci P, De Gioia L. Inorg Chem 2003;42:4773. [PubMed: 12870970]
146. Bruschi M, Fantucci P, De Gioia L. Inorg Chem 2002;41:1421. [PubMed: 11896710]
147. Fan H-J, Hall MB. J Am Chem Soc 2001;123:3828. [PubMed: 11457119]
148. Bruschi M, Fantucci P, De Gioia L. Inorg Chem 2004;43:3733. [PubMed: 15180430]
149. Tard C, Liu X, Ibrahim SK, Bruschi M, De Gioia L, Davies SC, Yang X, Wang LS, Sawers G, Pickett CJ. Nature 2005;433:610. [PubMed: 15703741]
150. Zampella G, Greco C, Fantucci P, De Gioia L. Inorg Chem 2006;45:4109. [PubMed: 16676972]
151. Boyke CA, van der Vlugt JI, Rauchfuss TB, Wilson SR, Zampella G, De Gioia L. J Am Chem Soc 2005;127:11010. [PubMed: 16076208]
152. Zampella G, Bruschi M, Fantucci P, De Gioia L. J Am Chem Soc 2005;127:13180. [PubMed: 16173745]
153. George SJ, Cui Z, Razavet M, Pickett CJ. Chem Eur J 2002;8:4037.
154. Zampella G, Bruschi M, Fantucci P, Razavet M, Pickett CJ, De Gioia L. Chem Eur J 2005;11:509.
155. Zampella G, Kravitz JY, Webster CE, Fantucci P, Hall MB, Carlson HA, Pecoraro VL, De Gioia L. Inorg Chem 2004;43:4127. [PubMed: 15236524]
156. Zampella G, Fantucci P, Pecoraro VL, De Gioia L. J Am Chem Soc 2005;127:953. [PubMed: 15656634]
157. Todeschini R, Moro G, Boggia R, Bonati L, Cosentino U, Lasagni M, Pitea D. Chemom Intell Lab Syst 1997;36:65.
158. Cramer RD, Patterson DE, Bunce JD. J Am Chem Soc 1998;110:5959.
159. Forloni G, Varì MR, Colombo L, Bugiani O, Tagliavini F, Salmona M. Curr Med Chem – Imun Endoc and Metab Agents 2003;3:185.

160. Forloni G, Iussich S, Awan T, Colombo L, Angeretti N, Girola L, Bertani I, Poli G, Caramelli M, Grazia Bruzzone M, Farina L, Limido L, Rossi G, Giaccone G, Ironside JW, Bugiani O, Salmona M, Tagliavini F. Proc Natl Acad Sci USA 2002;99:10849. [PubMed: 12149459]

161. Forloni G, Colombo L, Girola L, Tagliavini F, Salmona M. FEBS Lett 2001;487:404. [PubMed: 11163366]

162. Tagliavini F, Forloni G, Colombo L, Rossi G, Girola L, Canciani B, Angeretti N, Giampaolo L, Peressini E, Awan T, De Gioia L, Ragg E, Bugiani O, Salmona M. J Mol Biol 2000;300:1309. [PubMed: 10903871]

163. Chopra I, Roberts M. Microbiol Mol Biol Rev 2001;65:232. [PubMed: 11381101]

164. Mitscher, LA. The chemistry of tetracycline antibiotics. Marcel Dekker; New York: 1978.

165. De Leenheer AP, Nelis HJ. J Pharm Sci 1979;68:999. [PubMed: 480181]

166. Schneider, S. Tetracyclines in biology, chemistry and medicine. Greenwald, RA.; Hillen, W.; Nelson, M., editors. Birkhaüser Verlag; Basel: 2001. p. 65

167. Cosentino U, Varì MR, Saracino AAG, Pitea D, Moro G, Salmona M. J Mol Model 2005;11:17. [PubMed: 15592898]

168. Chang G, Guida WC, Still WC. J Am Chem Soc 1989;111:4379.

169. Saunders M, Houk KN, Dong Wu Y, Still CW, Lipton M, Chang G, Guida WC. J Am Chem Soc 1990;112:1419.

170. Still WC, Tempczyk A, Hawley RC, Hendrickson T. J Am Chem Soc 1990;112:6127.

171. Weiser J, Weiser AA, Shenkin PS, Still WC. J Comput Chem 1998;19:797.

172. Mohamadi F, Richards NGJ, Guida WC, Liskamp R, Lipton M, Caufield C, Chang G, Hendrikson T, Still WC. J Comput Chem 1990;11:440.

173. Maestro Molecular Modeling Interface V. 4.1.012. Schrödinger, Inc; Portland, Oregon 97201:

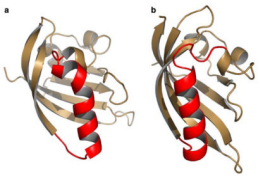

Fig. 1.

Cartoon representation of the mAhR LBD structure obtained by homology modeling, using: **a** the FixL template; **b** the HIF- 2α and ARNT templates. The different arrangements of the helical connector and the connecting loops are *highlighted*

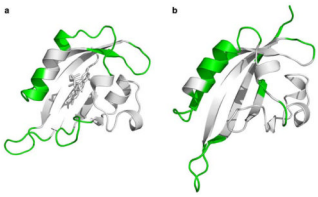

Fig. 2.

Locations of the motion patterns *highlighted* in the cartoon representation of the PAS domain structure of: **a** Phy3; **b** HIF-2α

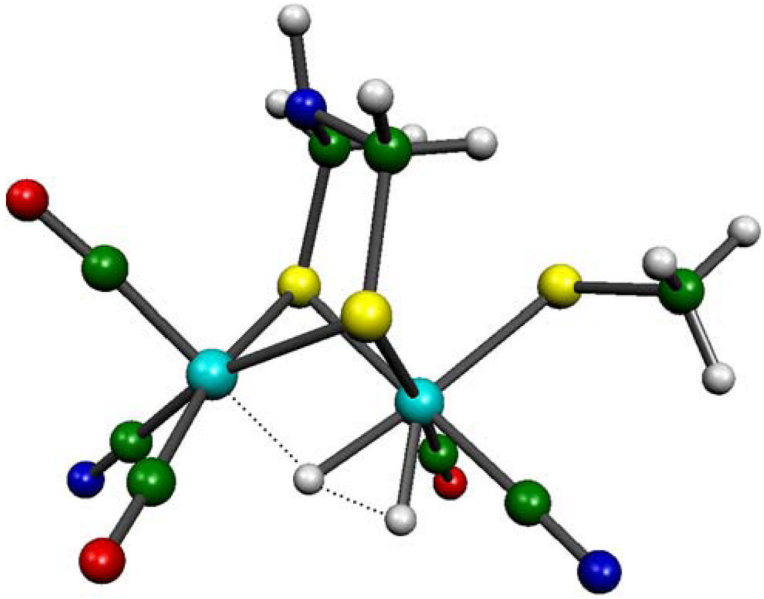

Fig. 3.

DFT (B-P86/TZVP on all atoms) optimized transition state structure of the di-hydrogen splitting reaction carried out by a model of the [Fe] only Hydrogenase cofactor. *Dashed line* shows the reaction coordinate. Atoms are coloured according to the following scheme: H, *white*; C, *green*; N, *blue*; Fe, *pale blue*; S, *yellow*

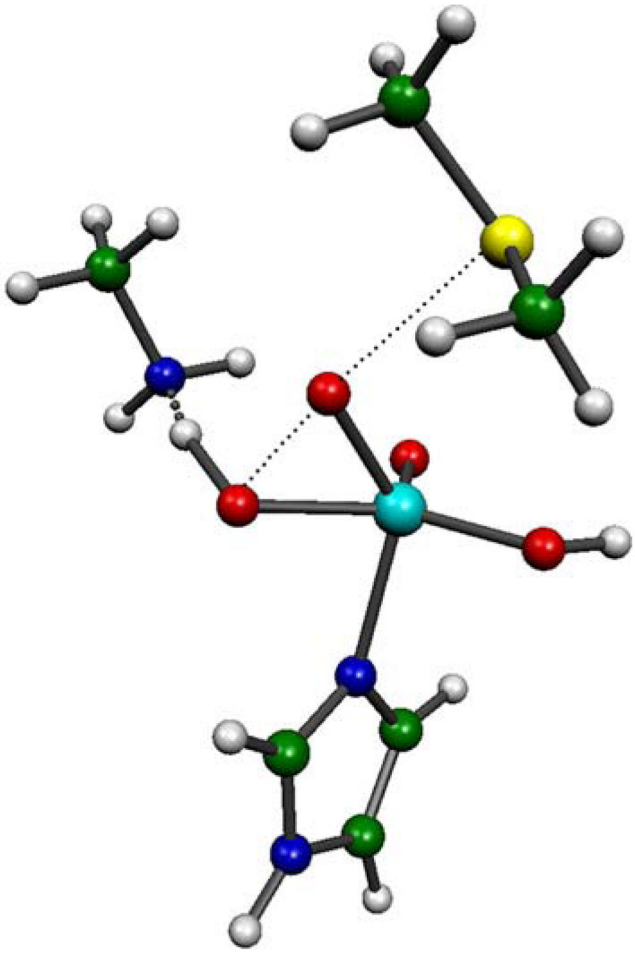

Fig. 4.

DFT (B-P86/TZVP on all atoms) optimized transition state structure of dimethyl sulphide oxidation performed by a model of the Vanadium Chloroperoxidase cofactor. A methyl-amine molecule (extreme left) was employed to model a lysine residue that is fundamental to turnover efficiency. *Dashed line* shows the reaction coordinate. Atoms are coloured according to the following scheme: H, *white*; C, *green*; N, *blue*; Fe, *pale blue*; S, *yellow*

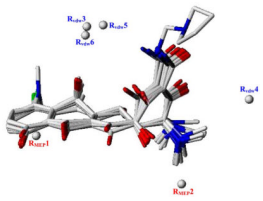

Fig. 5.
Descriptors entering the 3D-QSAR model of the defibrillogenic activity of tetracycline data set: spatial disposition around the tetracycline aligned structures