University of Florence

International Doctorate in Structural Biology

Cycle XVIII (2003-2006)

Theoretical studies on biomolecular systems: characterization, mobility and interactions

Ph.D. thesis of

Andrea Giachetti

Tutor

Prof. Ivano Bertini

Coordinator

Prof. Claudio Luchinat

SUMMARY

1. Introduction

1.1 Computational Chemistry in Structural Biology 4

2. Methods

2.1 Molecular Dynamics and NMR spin relaxation 7
2.1.1 Molecular Mechanics parameters 8
2.1.2 Molecular Dynamics simulation 10
2.1.3 Nuclear Magnetic Resonance spin relaxation 11
2.1.4 Model-free analysis 12

2.2 NMR protein structure calculation 14
2.2.1 NOEs derived restraint 15
2.2.2 Pseudocontact shift restraint 15
2.2.3 Residual dipolar coupling restraint 16
2.2.4 Relaxation rates derived restraint 17
2.2.5 Cross-correlation rates between Curie spin relaxation and dipole-dipole relaxation restraint 18

2.3 Introduction to Docking calculations 20
2.3.1 Molecular presentation of docking 20
2.3.2 Search methods and molecular flexibility 21
2.3.3 Scoring functions 21
2.3.4 Genetic algorithms 22

3. Publications

3.1 Modeling the Backbone Dynamics of Reduced and Oxidized Solvated Rat Microsomal Cytochrome b_5
[*Biophys. J.* 87, 498-512 (2004)] 25
3.2 Paramagnetics-based restraints for Xplor-NIH
[*J. Biomol. NMR* 28, 249-261 (2004)] 41
3.3 Combining in *silico tools* and NMR data to validate protein-ligand structural models: application to matrix metalloproteinases
[*J. Med. Chem.* in press] 55

4. Conclusions 104

5. Bibliography 106

1. Introduction

1.1 Computational chemistry in structural biology

Bioinformatics methods play a central role in the structural biology studies, and this field of research is fundamental for structure determination and mobility studies(*1*, *2*). Protein structure and mobility calculations are essential to understanding the structure–function relationship, and are indeed a major initiative in structural biology, which involves research in biochemistry, molecular biology, instrumental methods such as X-ray(*3*) crystallography and nuclear magnetic resonance (NMR) spectroscopy(*4*).

Molecular Dynamics simulations(*5*-*7*) of biomolecules can provide useful details concerning individual particle motion as a function of time. This can be used to address specific problems including properties of biological systems often not easily obtained from experimental measurements. Of course experiments play an essential role in validating the simulation methodology; comparison of simulation and experimental data serve to test the accuracy of the calculated results and to provide criteria for improving the methodology.

Computational modeling has been used to rationalize experimental data and to predict biophysical properties related to protein structure, dynamics and interactions(*8*). The three-dimensional structures determined using protein NMR spectroscopy provide an overall time-averaged or ensemble averaged topology, which is useful but does not fully represent the dynamic character of proteins or protein complexes. Often, interaction such as protein-protein or ligand-protein binding cannot be understood in terms of structures alone. Structural rearrangements and dynamic motions are typically involved in such intra- and inter-molecular interactions. Classical molecular dynamic (MD) computer simulations, based on experimentally determined structures, typically provide information for processes in the picosecond to microsecond time scale. Besides understanding the native motions of biomolecules, MD simulations have been used to understand biological processes, such as folding, conformational transitions, protein or

ligand binding, enzymatic catalysis, biological reactions and protein–membrane interactions(*9, 10*).

Molecular modeling and computational chemistry are assuming an increasingly important role in understanding the basis of drug-receptor interaction and assisting the medicinal chemist in the design of new therapeutic agents(*11-13*). The number of proteins with a known three-dimensional structure is increasing rapidly, and structures produced by structural genomics initiatives are beginning to become publicly available(*14*). The increase in the number of structural targets is in part due to improvements in techniques for structure determination, such as high-throughput X-ray crystallography(*15*) and NMR spectroscopy(*16, 17*). Due to the availability of such large amount of novel information the development of new theoretical approaches in docking programs has become a crucial part in pharmaceutical research and drug discovery.

Aims and topics of the research

The research carried out during my Doctorate in Structural Biology allowed me to acquire and develop techniques for structural characterization of biological systems, and for studies on mobility and possible interactions with other biomolecules.

The first work carried out was a theoretical investigation on the different mobility in cyt b_5 in both reduced and oxidized forms through molecular dynamics simulations(*18*). The order parameter (defined in the 'model free'(*19, 20*) theory) of ^{15}N nucleus of protein backbone was extracted from MD trajectories and compared with experimental NMR data(*21, 22*).

The second part of my work involved the implementation of a complete set of restraints based on paramagnetic(*23-25*) effects in the program Xplor-NIH(*26*), one of the most widely used programs for NMR structure determination. I optimized the "simulated annealing"(*27*) process to solve NMR structures of proteins using these restraints. The protocol was tested on different proteins for its validation(*28*).

In the last part of my work I developed a ligand-protein docking protocol able to take into account protein flexibility and experimental NMR restraints. This was done combining programs like AutoDock(*29*) with algorithms run with Xplor-NIH. AutoDock uses a

powerful genetic algorithm to search the lowest ligand-protein interaction energy, but cannot use experimental restraints and cannot admit protein flexibility. Xplor-NIH has a less powerful simulated annealing algorithm but can include experimental restraints and protein flexibility. This research was possible thanks to the experience acquired in my previous work, and based on the knowledge of MD and structure determination through experimental restraints techniques.

2. Methods

2.1 Molecular dynamics and NMR spin relaxation

Molecular dynamics simulations play a central role in the comprehension of biological systems. These simulations can provide important information on molecular motions and their impact on observable NMR. Experimental measurements of NMR spin relaxation also provides useful information on conformational fluctuations occurring in time scales from picoseconds to nanoseconds(*30*). Simulations on these time scales, has only recently become possible due to the improvements in computer power, and are likely to provide important new information on internal motions, overall diffusion and coupling between internal and rotational motions(*31*). Such information could provide a sound foundation for a detailed interpretation of NMR spin relaxation results. In principle, relaxation measurements can provide a valuable benchmark for judging the quantitative accuracy of simulations. For instance the different mobility of proteins in different oxidation states observed in NMR experiments can be simulated through molecular dynamics methods(*18*).

In order to simulate the structure and dynamics of biomolecules, they are approximated as a physical network of balls that have point charges at their centers and are connected by springs. In addition to these springs that govern the bending of bonds and angles, there are forces that favor certain rotations about the bonds. The balls representing the atoms are not hard spheres; they are Lennard-Jones particles that can overlap each other. Therefore MD simulations require thousands of potential energy parameters including springs stiffnes, equilibrium distances among atoms, torsional barriers periodicities, partial charges, and Lennard-Jones coefficients. The final goal is to study the motion of this physical network of balls and springs for the interpretation and prediction of dynamics of real macromolecules at the atomic level.

Given the initial coordinates of all atoms in a system we need to calculate the energy of the latter as a function of the atomic coordinates. Such function must be chosen together with the associated numerical constants. Both energy function and numerical protein energy parameters require extensive optimization.

2.1.1 Molecular Mechanics parameters

All interactions commonly employed in the potential energy function are summarized below. Simple harmonic terms describe bond stretching and angle bending. The planarity of groups (e.g., protein amide planes and aromatic rings) can also be enforced by harmonic potentials known as improper dihedrals. Rotation about single bonds (torsions) are governed by sinusoidal energies.

The bond stretching between two covalently bonded atoms i and j is represented by a harmonic potential:

$$V_s\left(r_{ij}\right) = \frac{1}{2} k_{ij}\left(r_{ij} - r_{ij}^0\right)^2 \qquad (2.1)$$

where r_{ij} is the bond distance, r_{ij}^0 is the equilibrium bond lengths and k_{ij} is bond stretching force constant. The bond angle vibration between a triplet of atoms i - j - k is also represented by a harmonic potential function :

$$V_\theta\left(\theta_{ijk}\right) = \frac{1}{2} k_{ijk}^\theta (\theta_{ijk} - \theta_{ijk}^0)^2 \qquad (2.2)$$

where θ_{ijk} is the angle between the vectors connecting the i, j atoms and j, k atoms respectively, θ_{ijk}^0 is the equilibrium angle value, k_{ijk}^θ are angle bending force constant.

Proper dihedral angles are defined according to the IUPAC/IUB convention, where Φ_{ijkl} is the angle between the ijk and the jkl planes, with zero corresponding to the *cis* configuration (i and l on the same side). The considered potential energy function is:

$$V_\phi(\phi_{ijkl}) = \frac{1}{2} k_\phi (1 + \cos(3\phi - \phi^0)) \qquad (2.3)$$

Improper dihedrals are meant to keep planarity groups planar (e.g., protein amide planes and aromatic rings) or to prevent molecules from flipping over to their mirror images. Their energy function is of the same type as that of angle bending potential.

The electrostatic attraction or repulsion between two charges is described by Coulomb's law:

$$V_{ij}^{Coulomb}(r_{ij}) = \frac{q_i q_j}{4\pi\varepsilon_r \varepsilon_0 r_{ij}} \qquad (2.4)$$

where q_i and q_j are the atoms' partial charges, r_{ij} is the distance separating the centers of atoms, ε_0 is the permittivity of free space, ε_r and is the relative dielectric coefficient of the medium between the charges.

van der Waals interactions must also be included. These are due to the electron cloud of a neutral atom fluctuating around the positively charged nucleus. The fluctuations in neighboring atoms become correlated, inducing attractive dipole-dipole interactions. The equilibrium distance between two proximal atomic centers is determined by a trade off between this attractive dispersion force and a core-repulsion force that reflects electrostatic repulsion. The Lennard Jones potential models the attractive interaction as $\alpha\ r^{-6}$ and the repulsive one as $\alpha\ r^{-12}$:

$$V_{ij}^{LJ}(r_{ij}) = \varepsilon_{ij}\left[\left(\frac{r_{ij}^0}{r_{ij}}\right)^{12} - 2\left(\frac{r_{ij}^0}{r_{ij}}\right)^{6}\right] \qquad (2.5)$$

where r^0 is the equilibrium separation distance (where the force $F = -dV_{ij}^{LJ} / dr_{ij} = 0$) and ε is the well depth ($V_{ij}^{LJ}(r_{ij}^0) = -\varepsilon_{ij}$)

A typical potential energy function used in MD simulations looks like:

$$V_{tot} = V_{tot}^{bonded} + V_{tot}^{nonbonded} \qquad (2.6)$$

$$V_{tot}^{bonded} = \sum_{bonds} V_s + \sum_{angles} V_\theta + \sum_{torsions} V_\phi + \sum_{impropers} V_\theta^{imp} \qquad (2.7)$$

$$V_{tot}^{nonbonded} = \sum_{i<j} \left(V^{LJ} + V^{Coulomb}\right) \qquad (2.8)$$

V_s is obtained as the sum bonds between atom pairs, V_θ is the sum over bond angles, V_ϕ and V_θ^{imp} are the sum over proper and improper torsion angles. In the nonbonded interactions (van der Waals and electrostatics), the summation is over atoms *i* and *j*, where '$i < j$' simply ensures that each interaction is counted only once. Generally, atoms separated by one or two bonds are excluded from the nonbonded sum, and those separated by three bonds, `1-4 interactions', may have electrostatic interactions reduced by a multiplicative scale factor. The form shown here reflects the choice not to include an explicit hydrogen bond term, favoring instead to account for hydrogen bonds through an appropriate parameterization of Lennard-Jones and Coulomb interactions.

2.1.2 Molecular dynamics simulation

MD simulation generally begins where experimental structure determination stops, if not during the structure refinement itself. It is generally not used to predict a structure directly from a sequence or to model the protein folding pathway. MD simulation can fold extended sequences to `global' potential energy minima for very small systems (peptides of length ten, or so, in vacuum), but it is most commonly used to simulate the dynamics of known structures. An initial velocity is assigned to each atom, and Newton's laws are applied at the atomic level to propagate the system's motion through time. Thus, dynamical properties such as time correlation functions and transport coefficients (e.g., diffusion constants, bulk viscosities) can be calculated from a sufficiently long MD trajectory.

Classical MD algorithm is derived from Newton's second law $F_i = m_i a_i$, where F_i is the sum of all forces acting on atom i, that results in its acceleration. The acceleration is the second derivative of the position with respect to time: $a_i = dv_i/dt_i = d^2 r_i/dt^2$.

The `Leap Frog' algorithm is one method commonly used to numerically integrate Newton's second law. We obtain all atomic positions r_i at all times t and all atomic velocities v at intermediate times $t_{n+1/2}$. This method gets its name from the way in which positions and velocities are calculated in an alternating sequence, `leaping' past each other in time:

$$\vec{v}_i(t_{n+1/2}) = \vec{v}_i(t_{n-1/2}) + \frac{\vec{F}_i(t_n)}{m_i}\Delta t \qquad (2.9)$$

$$\vec{r}_i(t_{n+1}) = \vec{r}_i(t_n) + \vec{v}_i(t_{n+1/2})\Delta t \qquad (2.10)$$

Initial velocities are assigned so as to reflect equilibrium at the desired temperature *T* (a Maxwell distribution), without introducing a net translation or rotation of the system. The energy of an isolated system (as opposed to, for example, one in contact with a thermal bath) is conserved in nature, but it may not be in simulations. Energy conservation can be violated in simulations because of an insufficiently short integration time step Δt, an inadequate cutoff method applied to long range (electrostatic and Lennard-Jones) forces, or even bugs in the program. Of course, energy conservation alone is not sufficient to ensure a realistic simulation. The realism of the dynamics trajectory depends on the empirical potential energy function V_{tot}, the treatment of long-range forces as well as the value of Δt .

2.1.3 Nuclear Magnetic Resonance spin relaxation

Nuclear magnetic resonance relaxation measurements provide an important probe of the dynamics of molecules. The spin-lattice (T_1) and spin-spin (T_2) relaxation times and NOE enhancement factor of protonated heteronuclei are a function of the thermal motions, and relaxation is dominated by the dipolar interaction with the directly attached ^{1}H spin and by the chemical shift anisotropy mechanism(*32*, *33*). The relaxation constants are determined by the spectral density functions *J(ω)*, according to,

$$\frac{1}{T_1} = \frac{d^2}{4}\left[J(\omega_H - \omega_X) + 3J(\omega_X) + 6J(\omega_H + \omega_X)\right] + c^2 J(\omega_X) \quad (2.11)$$

$$\frac{1}{T_2} = \frac{d^2}{8}\left[4J(0) + J(\omega_H - \omega_X) + 3J(\omega_X) + 6J(\omega_H) + 6J(\omega_H + \omega_X)\right] + \frac{c^2}{6}\left[3J(\omega_X) + 4J(0)\right] + R_{ex} \qquad (2.12)$$

$$NOE = 1 + \frac{1}{R_1}\frac{\gamma_H}{\gamma_X}\frac{d^2}{4}\left[6J(\omega_H + \omega_X) - J(\omega_H - \omega_X)\right] \qquad (2.13)$$

where $c = \omega_X \Delta\sigma / \sqrt{3}$, $d = \mu_\circ h \gamma_H \gamma_X r_{HX}^{-3} / 8\pi^2$, μ_0 is the permeability of free space; *h* is Planck's constant; γ_H and γ_H are the gyromagnetic ratios of ^{1}H and X spin (X=^{13}C or

^{15}N),respectively; r_{XH} is the X-H bond length; ω_H and ω_X are the Larmor frequencies of ^{1}H and X spin; and $\Delta\sigma$ is the chemical shift anisotropy of the X spin (assuming an axially symmetric chemical shift tensor). The symmetry axis of the chemical shift tensor is assumed to be collinear with the X-H bond vector. R_{ex} is introduced to account for other processes that contribute to $1/T_2$. Usually a non-zero R_{ex} implies the presence of motions on the microsecond to millisecond time scale. For typical magnetic fields available at present, ω_N and ω_H range from 50 MHz up to 900 MHz.

The spectral density function is the Fourier transform of the angular auto-correlation function, *C(t)*, of the N-H bond vector,

$$j(\omega) = 2\int_0^\infty C(t)\cos\omega t dt\,. \qquad (2.14)$$

The correlation function describes reorientation of the N-H bond vector due to both the overall molecular tumbling and internal dynamics. Assuming that the overall molecular tumbling is much slower than the internal motions, one can decompose *C(t)* as

$$C(t) = C_0(t)C_I(t) \qquad (2.15)$$

where $C_0(t)$ and $C_i(t)$ are the correlation functions for the overall tumbling and internal motions respectively. When the molecular tumbling is isotropic, the overall motion can be described by a single correlation time, $\tau_M^{-1} = 6D_m$

$$C_0^{iso}(t) = \frac{1}{5}e^{-t/\tau_M} \qquad (2.16)$$

where D_M is the isotropic rotational diffusion constant. The internal correlation function is given as

$$C_I(t) = \left\langle P_2(\hat{\mu}(0))\cdot\hat{\mu}(t))\right\rangle \qquad (2.17)$$

where the second Legendre polynomial $P_2(x) = (3x^2 - 1)/2$, and the unit vector μ describes the orientation of the N-H bond vector in the molecular reference frame.

2.1.4 Model-free analysis

As shown in equations (2.11,2.12,2.13), the relaxation constants are determined by the spectral densities at five characteristic frequencies. The spectral density mapping can be

rebuild using the so-called model-free formalism(*19, 20*). The unknown spectral density function, or equivalently, the internal correlation function, is modeled by simple analytical functions that depend on a few 'model-free' parameters. For example, the internal dynamics can be simply characterized by two motional parameters,

$$C_I(t) = S^2 + (1 - S^2)e^{-t/\tau_e} \qquad (2.18)$$

in which the squared generalized order parameter, $S^2 = C_I(\infty)$, reflects the amplitude of the internal motions, which time constant, τ_e, in this case, equals to the integrated correlation time, τ_{eff}, defined as

$$\tau_{eff} = \frac{1}{C_I(0) - C_I(\infty)} \int_0^\infty (C_I(t) - C_I(\infty))dt \qquad (2.19)$$

The angular reorientational correlation function that describes the dynamics of an N-H bond can be calculated from MD simulation of length T using(*34*)

$$\left\langle P_2(\bar{\mu}(0) \cdot \bar{\mu}(t)) \right\rangle = \frac{1}{T - t} \int_0^{T-t} P_2(\bar{\mu}(\tau) \cdot \bar{\mu}(t + \tau)d\tau \qquad (2.20)$$

The long time of this function is defined as the square of the generalized order parameter (S^2) and can be obtained from the simulation

$$S^2 = \frac{1}{T^2} \int_0^T \int_0^T P_2(\bar{\mu}(\tau) \cdot \bar{\mu}(t + \tau)d\tau dt \qquad (2.21)$$

For trajectory of finite length that do not adequately sample conformation space, the long time limit calculated using equation (2.20) will in general differ from the value obtained from equation (2.21) In order to remove the overall motion contribution, all trajectory snapshots have to be superimposed.

2.2 NMR protein structure calculation

NMR spectroscopy is a technique widely used for protein structure determination. It allows obtaining structures of proteins in solution, i.e., in environments much closer to the natural biological system than the single crystals required for protein crystallography, with similar accuracy to the latter. Geometric conformational restraints have to be derived from available experimental NMR data, in order to be used as structural restraints(*35*). These restraints should convey to the calculated structure all the information present in the NMR data. Experimental NMR data alone however are not sufficient to determine the position of all atoms in a biological macromolecule. It has to be supplemented by information about the covalent structure of the protein (the amino acid sequence, bond lengths, bond angles, chiralities, and planar groups) and the steric repulsion between nonbonded atom pairs. Depending on the structure calculation programs, the structure parameters and the conformational restraints that are employed can be different. The most widely used programs in NMR structure determination are DYANA(*36*), CYANA(*37*), CNS(*38*) and Xplor-NIH. All programs perform MD simulations in cartesian or torsional angle space.

The purpose of a molecular dynamics calculation in an NMR structure determination program is that of searching the conformation space of the protein for structures that fulfill the restraints and minimize a target function defined as the sum of the potential energies of the system (equation 2.6) and of the energy contributions related to the different classes of available restraints. The minimization of the target function defined is found using a simulated annealing procedure. This is a special case of MD simulation performed initially at high temperature using a simplified force field that treats the atoms as soft spheres without attractive or long-range (i. e. electrostatic) nonbonded interactions and does not include explicit consideration of the solvent. The temperature is then gradually reduced during the simulation. Often, the system is first heated and then cooled. In this way the system has the possibility to cross energetic barriers. Of course, annealing is more time expensive than energy minimization. The energy functions that is minimized includes unphysical energy terms as for instance those related to NMR derived restraints.

2.2.1 NOEs derived restraint

The nuclear Overhauser effect(*39*) provides information on the spatial proximity of nuclei. NOE determinations are usually homonuclear, in the case of protons, but can also be heteronuclear, with ^{1}H signals irradiated and those of heteronuclei observed. Use of NOEs play the primary role for three-dimensional structure determination. The intensity of NOE is proportional to the inverse of the sixth power of the distance between the two nuclei observed. Usually NOEs are observed only for proton pairs separated by less than 6-7 Å. In the structure determination algorithm, NOE distance restraints are used as upper bound limits, because NOEs may be reduced by internal motions and exchange effects.

2.2.2 Pseudocontact shift restraint

In paramagnetic systems, the NMR lines are affected by the unpaired electrons that generate an extra contribution to the chemical shift, which is called hyperfine shift. The different values of the observed chemical shift between a paramagnetic system and a diamagnetic analog is the sum of the pseudocontact shift contribution δ^{pcs} and the contact shift contribution δ^{con}. The contact shift contribution is correlated to the amount of unpaired spin density of the resonating nucleus(*24*):

$$\delta^{con} = \frac{A_c}{\hbar}\frac{g_e \mu_B S(S+1)}{3\gamma_N kT} \qquad (2.22)$$

where A_c is the nucleus-electron hyperfine coupling constant, g_e is the free electron g factor, μ_B is the Bohr magneton, S is the electron spin moment, γ_N is the nuclear magnetogyric ratio, $\hbar$ is reduced Planck constant, k is Boltzmann constant, T is the temperature.

The pseudocontact shift arises from the magnetic susceptibility anisotropy, and is given by(*40-43*):

$$\delta^{pcs} = \frac{1}{12\pi r^3}\left[\Delta\chi_{ax}(3\cos^2\theta - 1) + 3/2\,\Delta\chi_{rh}\sin^2\theta\cos 2\varphi\right] \qquad (2.23)$$

$$\Delta\chi_{ax} = \chi_{zz} - \frac{\chi_{xx} + \chi_{yy}}{2} \text{ and } \Delta\chi_{rh} = \chi_{xx} - \chi_{yy} \qquad (2.24)$$

where θ and φ identify the polar coordinates of the nucleus in the frame of the magnetic field susceptibility tensor, and $\Delta\chi_{ax}$ and $\Delta\chi_{rh}$ are the axial and rhombic anisotropy parameters of paramagnetic susceptibility tensor of the metal. Pseudocontact shift restraints alone cannot be used to solve the protein three-dimensional structure, but when used together with NOEs and dihedral angle restraints thay contain precious information useful for protein refinement and for ab-initio structural calculations(*44-53*). Therefore they were implemented into the Xplor-NIH package.

2.2.3 Residual dipolar coupling restraint

Residual dipolar couplings, induced by alignment of the molecules with the external magnetic field B_0, can provide important long-range restraints for structure determination(*54*). The presence of a paramagnetic center induces different energy values for the orientations of the protein related to the magnetic susceptibility tensor. The presence of this sizeable anisotropy in the latter tensor leads the average nucleus-nucleus dipolar coupling to a value different from zero. A small anisotropy of the magnetic susceptibility tensor is also present in diamagnetic proteins but is often smaller than that of paramagnetic systems. Due to the diamagnetic contribution, the magnetic susceptibility tensor anisotropies estimated from residual dipolar couplings are different from those calculated from pseudocontact shifts. RDC values can be derived from the difference between the corresponding $^1J_{\mathrm{HN}}$ values for the paramagnetic and the diamagnetic species, measured at the same field. In this case only the paramagnetic anisotropy is responsible of the experimentally obtained RDCs.

The expression for the residual dipolar coupling between two directly coupled nuclei can be simplified to the form(*42, 55*)

$$rdc = \frac{1}{4\pi}\frac{B_0^2}{15kT}\frac{\gamma_I\gamma_S\hbar}{2\pi r_{IS}^3}\left[\Delta\chi_{ax}(3\cos^2\theta - 1) + 3/2\,\Delta\chi_{rh}\sin^2\theta\cos 2\varphi\right] \quad (2.24)$$

where $\Delta\chi_{ax}$ and $\Delta\chi_{rh}$ are the axial and rhombic anisotropies of the molecular magnetic susceptibility tensor, θ and φ are polar coordinates describing the orientation of the I-S bond vector in the axis system of the tensor, r_{IS} is the I-S distance, γ_I is the nuclear magnetogyric ratio of nucleus I, γ_S is the nuclear magnetogyric ratio of nucleus S, $\hbar$ is reduced Planck constant, k is Boltzmann constant, T is the temperature.

The long-range information content of the residual dipolar couplings is quite different from the short-range nature of the restraints derived from NOEs and pseudocontact shifts, and can significantly improve the quality of macromolecular NMR structures(*56-58*). Therefore, also paramagnetic RDC have been included in Xplor-NIH program (see appendix 4) .

2.2.4 Relaxation rates derived restraint

In paramagnetic proteins Nuclear relaxation rates (R_1 and R_2)of resonating nuclei are affected by the presence of dipolar coupling with the unpaired electron. The paramagnetic contribution to longitudinal dipolar relaxation rate R_{1M} is described by the equation(*25*):

$$R_{1M} = \frac{k}{r^6} \tag{2.25}$$

where r is the distance between paramagnetic center and resonating nuclei and k is a constant

$$k = \frac{2}{15}\left(\frac{\mu_0}{4\pi}\right)^2 \gamma_I^2 \mu_{eff}^2 \left[\frac{7\tau_c}{1+\omega_s^2\tau_c^2} + \frac{3\tau_c}{1+\omega_I^2\tau_c^2}\right] \tag{2.26}$$

$$\tau_c^{-1} = \tau_r^{-1} + \tau_M^{-1} + \tau_s^{-1} \tag{2.27}$$

where τ_c depend from the rotational correlation rate τ_r, the exchange rate τ_m and the electron relaxation rate τ_s, $\mu_{eff}^2 = g_e^2 \mu_B^2 S(S+1)$ or $\mu_{eff}^2 = g_J^2 \mu_B^2 J(J+1)$ for lanthanides and actinides, γ_I is proton magnetogyric ratio, ω_I and ω_S are proton and electron Larmor frequency. The relaxation rate depends on the inverse of the 6th power of the metal-nucleus distance. R_1 value are experimentally available in a spherical shell where the effect is not too weak to be detected and not too strong to increase the line width so that the signal itself is not observable any more. The paramagnetic contribution of nuclear relaxation is used as upper distance limit restraint and it is obtained as difference between the longitudinal relaxation rate of paramagnetic and diamagnetic samples(*59*). Implementation of these restraints in Xplor-NIH is been part of the present work.

2.2.5 Cross-correlation rates between Curie spin relaxation and dipole-dipole relaxation restrain

The paramagnetic center induces relaxation of nuclear spins due to the presence of dipolar interaction between the static electron magnetic moment and the magnetic field. This interaction depends on the relative populations of the electronic energy levels. This contribution to relaxation is called Curie spin relaxation, and is modulated by the rotational correlation time of the molecule(*60, 61*).

Curie spin relaxation and dipole-dipole nuclear relaxation have the same correlation time (the rotational correlation time τ_r). This causes the occurrence of cross correlation effects between the two mechanisms, which give rise to differential line broadening of the doublet components, α and β, in a coupled two-spin system(*62*). This can be written as

$$R_2^{\alpha,\beta} = R_2^{all} \pm R_2^{CS-DD} \qquad (2.28)$$

where R_2^{CS-DD} is the cross correlation contribution to R_2, and R_2^{all} is the sum of all the other contributions. Measurements of R_2^{CS-DD} values provide structural information in macromolecules. In fact they are related to the distance, r_{IM}, between the nucleus I and the paramagnetic centre M, and to the angle θ between the I-M direction and the I-J direction (J being the nucleus coupled to I) by the following equation, which is valid for an isotropically tumbling molecule with an isotropic susceptibility tensor(*62*):

$$R_2^{CS-DD} = \frac{1}{10}\left(\frac{\mu_0}{4\pi}\right)^2 \frac{\hbar\gamma_I^2\gamma_J B_0 g_e^2 \mu_B^2 S(S+1)}{(3kT) r_{IM}^3 r_{IJ}^3}\left(3\cos^2\theta - 1\right)\left(4\tau_r + \frac{3\tau_r}{1+\omega_I^2\tau_r^2}\right), \qquad (2.29)$$

r_{IJ} is the distance between nuclei I and J with magnetogyric ratios γ_I and γ_J, respectively, S is the electron spin quantum number, τ_r is the rotational correlation time, ω_I is the precession frequency for the nuclear spin, and all the other symbols have their usual meaning. Note, for lanthanide ions, electron spin S and g_e in equation 2.29, have to be replaced with the total angular momentum quantum number J and g_J respectively.

The quantity that is experimentally measured is the difference in R_2 values from the two doublet lines (i.e. $R_2^{\alpha} - R_2^{\beta}$) indicated with Γ. Cross correlation between chemical shift anisotropy and dipolar relaxation also affects the observed values of Γ, and must be removed by subtracting the corresponding values measured on a diamagnetic

analogue(*63*). Implementation of these restraints in Xplor-NIH is part of the present work.

2.3 Introduction to docking calculations

Computational methodologies have become a crucial component of many drug discovery programs, from hit identification to lead optimization and beyond, and approaches such as ligand- or structure-based(*64*) virtual screening techniques are widely used in many discovery efforts(*13*). When only the structure of a target and its binding site are available, high-throughput docking is primarily used as a hit-identification tool. However, similar calculations are often used also later during lead optimization, in order to study modifications of the enzyme structure through computer models before preceding with compound synthesis. The availability of experimental restraints and a protocol to evaluate them would be very useful in order to calculate a correct protein-ligand conformation.

The docking process involves the prediction of ligand conformation and orientation (or posing) within a targeted binding site. The process begins with the application of docking algorithms that pose small molecules in the binding site. This in itself is challenging, as even relatively simple organic molecules can contain many conformational degrees of freedom. Sampling these degrees of freedom must be performed with sufficient accuracy to identify the conformation that best matches the receptor structure, and must be fast enough to permit the evaluation of thousands of compounds in a given docking run(*65*). The inclusion of few experimental restraints (like protein-ligand NOEs) in this calculation can reduce drastically the conformational space search and improve the accuracy of the calculation.

Algorithms are complemented by scoring functions that are designed to predict the biological activity through the evaluation of interactions between compounds and potential targets.

2.3.1 Molecular presentation of docking

To evaluate various docking methods, it is important to consider how protein and ligand are represented. There are three basic representations of the receptor: atomic, surface and grid(*66*). Among these, atomic representation is generally only used in standard molecular mechanics potential energy function(*67*) (as in Xplor-NIH see MD simulation

paragraph) Surface-based docking programs are typically, but not exclusively, used in protein–protein docking(*68*). These methods attempt to align points on surfaces by minimizing the angle between the surfaces of opposing molecules(*69*). Therefore, a rigid body approximation is still the standard for many protein–protein docking techniques. The basic idea of grid approach is to store information about the receptor's energetic contributions on grid points so that it only needs to be read during ligand scoring(*29*). In the most basic form, grid points store two types of potentials: electrostatic and van der Waals. The final energy is calculated from the sum of the interaction energy of any atom of the ligand (atomic representation) and every grid point of the receptor.

2.3.2 Search methods and molecular flexibility

Ligand flexibility can be divided into two basic categories: random or stochastic methods (Monte Carlo, genetic algorithms), and simulation methods (molecular dynamics, energy minimization).

Molecular dynamics is currently the most popular simulation approach. However, molecular dynamics simulations are often unable to cross high-energy barriers within feasible simulation time periods, and therefore might only accommodate ligands in local minima of the energy surface(*65*). The classical MD approach used to simulate protein–ligand system interaction is the simulated annealing method (*27*).

Random search algorithms operate by making random changes to either a single ligand or a population of ligands. A newly obtained ligand is evaluated on the basis of a pre-defined probability function. Two popular random approaches are genetic algorithms(*29, 70, 71*) and Monte Carlo search method(*72*). Genetic algorithms are used in several program including AutoDock (from 3.05 version), GOLD(*73*) and DOCK(*74*), Monte Carlo approach was used in AutoDock version 2.0.

2.3.3 Scoring functions

The evaluation and ranking of predicted ligand conformations is a crucial aspect of structure-based virtual screening. Even when binding conformations are correctly predicted, the calculations ultimately do not succeed if they do not differentiate correct poses from incorrect ones, and if 'true' ligands cannot be identified. Scoring functions

implemented in docking programs make various assumptions and simplifications in the evaluation of modeled complexes and do not fully account for a number of physical phenomena that determine molecular recognition — for example, entropic effects.

Most scoring functions used in different docking and scoring programs (AutoDock, GOLD, Gscore(*75*)) are based on molecular mechanics force-field(*76*). This approach usually quantify the sum of two energies, the receptor–ligand interaction energy and internal ligand energy. Interactions between ligand and receptor are most often described by using van der Waals and electrostatic energy terms. The van der Waals energy term is given by a Lennard–Jones potential function. Electrostatic terms are accounted for by a Coulombic formulation with a distance-dependent dielectric function that lessens the contribution from charge–charge interactions. The functional form of the internal ligand energy is typically very similar to the protein–ligand interaction energy, and also includes van der Waals contributions and/or electrostatic terms.

Standard force-field scoring functions have major limitations, because they were originally formulated to model enthalpic gas-phase contributions to structure and energetics, and do not include solvation and entropic terms.

AutoDock extensions of force-field-based scoring functions include a torsional entropy term for ligands, adding 0.3113 Kcal/mol for any torsional angles present in the ligand, and the inclusion of explicit protein–ligand hydrogen-bonding by a 12–10 Lennard–Jones potential with a directional component.

2.3.4 Genetic algorithms

For a protein-receptor with known three-dimensional structure, the ligand-protein docking problem basically consists of predicting the bound conformation of a ligand molecule within the protein active site. The docking problem is a difficult optimization problem involving many degrees of freedom, and the development of efficient docking algorithms and methodologies would be of enormous benefit in the design of new drugs. Genetic Algorithms(*77, 78*) are inspired in Darwin's theory of evolution by natural selection and are powerful tools in difficult search and optimization problems. The GA works with a population of individuals where each individual represents a possible solution for the problem to be solved and, in ligand-protein docking problem, it is the

position of the ligand with respect to the protein. Therefore, a ligand conformation is represented by a chromosome constituted by real valued genes representing ligand translational, orientational and conformational degrees of freedom. The individuals are evaluated by a fitness function, that is, the total interaction energy between the protein and the ligand molecule and the intramolecular ligand energy. Individuals in the population are selected for reproduction in accordance with their fitness, and undergo mutation and crossover reproduction operators, to generate new individuals.

The main idea of genetic algorithms was borrowed from genetics and natural selection. A population of chromosomes encoding solutions to the problem is first generated and then it evolves through a process similar to biological evolution. Chromosomes encoding lower ligand-protein energy interaction survive, reproduce and combine to generate new chromosomes, which hopeful encode better solutions in the succeeding generations. Chromosomes with small fitness will gradually perish in the succeeding generations.

The strength of genetic algorithms lies in its ability to handle a large and diverse set of variables, and have been considered more efficient then molecular dynamics methods like simulated annealing.

3. Publications

3.1 Modeling the Backbone Dynamics of Reduced and Oxidized Solvated Rat Microsomal Cytochrome b_5

[*Biophys. J.* 87, 498-512 (2004)]

Modeling the Backbone Dynamics of Reduced and Oxidized Solvated Rat Microsomal Cytochrome b_5

Andrea Giachetti,* Giovanni La Penna,*† Angelo Perico,† and Lucia Banci*
*Magnetic Resonance Center, University of Florence, Florence, Italy; and †National Research Council, Institute for Macromolecular Studies, Genoa, Italy

ABSTRACT In this article, a description of the statistics and dynamics of cytochrome b_5 in both reduced and oxidized forms is given. Results of molecular dynamics computer simulations in the explicit solvent have been combined with mode-coupling diffusion models including and neglecting the molecule-solvent correlations. R_1 and $R_{1\rho}$ nuclear magnetic relaxation parameters of ^{15}N in the protein backbone have been calculated and compared with experiments. Slight changes in charge density in the heme upon oxidation produces a cascade of changes in charge distributions from heme propionates up to charged residues ~1.5 nm from Fe. These changes in charge distributions modify the molecular surface and the water shell surrounding the protein. The statistical changes upon oxidation can be included in diffusive models that physically explain the upper and lower limits of $R_{1\rho}$ relaxation parameters at high off-resonance fields.

INTRODUCTION

Electron transfer (ET, hereafter) proteins are involved in a variety of important biological processes. These proteins exist in different oxidation states and, in most cases, they contain one or more metal ions which are donor or acceptor of electrons (Bertini et al., 2001, 1994; Kadish et al., 1999; Scott and Mauk, 1996). The different structural and dynamical properties of the two oxidation states have been the subject of a huge number of studies. In particular, the mobility of the redox forms of these proteins might be different and this difference can be related to protein function (Banci et al., 1998b; Brooks et al., 1988; McCammon and Harvey, 1987), molecular recognition and reorganization energy (Canters and Van de Kamp, 1992; Corin et al., 1991; Hake et al., 1992; Peerey et al., 1991; Walker and Tollin, 1991; Zhou and Kostic, 1992). Experimental, theoretical and computational tools are all required to monitor and understand the structural and mobility differences between redox forms at an atomic level.

Cytochrome b_5 (Cyt b_5, hereafter) is a largely widespread ET protein found in a variety of mammalian and avian species, and it has been extensively studied (Kadish et al., 1999; Lederer, 1994; Mathews, 1985). The protein is mainly found as a membrane-bound protein. However, a hydrophilic domain was identified and isolated, and it still retains its activity (Ito and Sato, 1968; Spatz and Strittmatter, 1971; von Bodman et al., 1986). Because of the moderate dimensions and availability, this domain represents an ideal system for addressing in vitro problems related to how the ET process occurs and how energy is transferred and stored in living organisms (Canters and Van de Kamp, 1992; Moore et al., 1986; Moore and Pettigrew, 1990; Scott and Mauk, 1996). Structures have been determined for some isoforms through x-ray only for the oxidized form, whereas solution structures of both oxidized (Arnesano et al., 1998a) and reduced (Banci et al., 1997a; Dangi et al., 1998b) forms of rat microsomal Cyt b_5 have become recently available, in addition to the crystal structure of the oxidized bovine liver protein (Durley and Mathews, 1996).

The comparison of the solution structures of Cyt b_5 in the two oxidation states revealed small differences, indicating rearrangements that could be relevant in determining the stability of the protein in the two redox states as well as in the ET process. Nevertheless, the structural differences were not significant enough to produce differences in physicochemical properties like the solvent accessibility, backbone mobility, and solvation entropy (Banci and Presenti, 2000; Dangi et al., 1998a,b).

Local mobility has often been thought to be important in controlling ET processes. Heteronuclear spin relaxation experiments, measured through nuclear magnetic resonance (NMR) spectroscopy, can provide direct information on local mobility (Palmer, 2001; Wagner et al., 1993; Wagner, 1993). In particular, rotating frame relaxation experiments have been used to reveal motions with characteristic times in the range of micro to milliseconds (Desvaux et al., 1995; Palmer et al., 2001). However, the interpretation of these NMR data is complicated by the high degree of cooperativity of molecular segments' motions in the relevant relaxation modes.

Two kinds of effects are expected with protein oxidation change. First, the change in electron structure, localized on the metal ion, affects the metal coordination geometry and the active site structure. Second, the change in the charge density in the active site modifies both the structure and mobility of

Submitted October 31, 2003, and accepted for publication March 4, 2004.

Address reprint requests to Lucia Banci, Magnetic Resonance Center, University of Florence, via L. Sacconi 6, 50019 Sesto Fiorentino (Florence), Italy. E-mail: banci@cerm.unifi.it.

0006-3495/04/07/498/15 $2.00

doi: 10.1529/biophysj.103.036657

the protein matrix and, eventually, of the protein-solvent interface. Models that are able to include in a unique frame local motions, as they are contained in the high frequency relaxation data and in the rotating frame data, are therefore required to connect molecular dynamics to NMR experiments. Computer simulations, and especially molecular dynamics (MD) simulations (Brooks et al., 1988; Kollman and Merz, 1990; van Gunsteren, 1993), allow calculations of NMR relaxation parameters taking into account the necessary structural atomic details of biological macromolecules (Case, 2002). Up to date, only a few MD and stochastic simulations are long enough to give a good statistical description of molecular rotational tumbling (Peter et al., 2001; Stocker and van Gunsteren, 2000). Usually, the measured NMR relaxation rates are analyzed within a model-free approach (Clore et al., 1990; Lipari and Szabo, 1982) which describes the involved time correlation functions by separating overall rotation and internal motions. The parameters involved in the fitting procedure, restricted or extended according to the sophistication of the fitting equations, are essentially order parameters and rotational and internal correlation times.

Prompers and Brüschweiler (2002) introduced a different approach to the interpretation of the NMR dynamics of proteins, named, in its more advanced form, isotropic reorientational eigenmode dynamics. This method gives first a statistical estimate of the isotropically averaged covariance matrix of the second-rank orientations of the network of relaxing vectors. Then, it derives the dynamics from an MD simulation by describing the time evolution of the projections, onto the static eigenvectors of the covariance matrix, of the second-rank components related to the relaxing vectors. The resulting approximate dynamics normally produces an overestimate of the relaxation rates that are more relevant in determining the NMR relaxation parameters. This problem is overcome by a fitting to the NMR relaxation data to decrease the lower relaxation rates.

A more basic approach couples the diffusion theory, to describe the time evolution of protein configuration, with computer simulations, used to estimate the time-independent statistics. This method has been applied both on MD (Fausti et al., 2000; La Penna et al., 2003; Perico and Pratolongo, 1997) and Monte Carlo (La Penna et al., 2003) simulations. In this article, we report a model of nuclear spin relaxation parameters of Cyt b_5 obtained by combining detailed MD computer simulations of the protein in the explicit solvent with diffusion theory. This tool allows the calculation of time correlation functions relevant to NMR relaxation without the need of the information on time evolution of configurational variables contained in the MD computer simulation. The structural and statistical properties obtained by MD are summarized and the application of diffusion theory to the calculation of molecular dynamics are presented. The theory provides a model for the upper and lower limits of the relaxation parameters, based on the statistical results of the computer simulation of the atomic model and without the need of parameter fitting or a priori assumptions on timescale separation.

The biological relevance of these findings as well as perspectives of this approach are discussed.

METHODS

Molecular dynamics simulation setup

The calculations to obtain the point charges and the MD statistics of Cyt b_5 in explicit water have been performed with the package NWCHEM (HPCC group, 2002). In protoporphyrin IX, or heme type b, only the Fe ion is covalently bonded to the protein. In Cyt b_5, the heme can be bonded in two different orientations related by a 180° rotation around the CHC-CHA *meso* direction. This possibility gives rise to two protein conformations called *A* and *B*, the ratio of which depends on the origin of Cyt b_5, being that conformation A is always the most populated.

The point charges of the heme, the Fe-bonded histidines and the iron ion (Fe(II) and Fe(III) in the reduced and oxidized states, respectively), have been computed ab initio by using the RESP procedure (Bayly et al., 2002) as it has been implemented in the NWCHEM package. For the oxidized form, the Hartree-Fock calculations have been restricted to a single unpaired electron, as it is experimentally found for the Fe^{+3} ion in this complex.

First, a 6-31G basis-set has been used to optimize the geometry of the heme only, where all methyl, vinyl, and propionate substituents in protoporphyrin IX have been replaced with hydrogen atoms. The initial D_{4h} symmetry of the heme group has not been imposed (Rovira et al., 2001). Second, two imidazole molecules have been added to the resulting heme configurations with the imidazole planes translated and oriented as in the Cyt b_5 crystallographic structure of recombinant trypsin-solubilized fragment of Cyt b_5 (PDB entry 1EHB) (Wu et al., 2000). In the A conformation, the imidazole planes are almost contained in the plane formed by the two segments CHB-CHD of heme and Nϵ2(ligand 1)–Nϵ2(ligand 2).

A detailed study of the electron energy obtained with Hartree-Fock calculations using the 6-31G* basis-set has been performed. By changing the orientation of one or two imidazole planes with respect to the CHB-CHD segment, several local minima have been identified: the two imidazole planes can be parallel to each other and parallel either to the C_{meso}-Fe-C_{meso} line or the N(heme)-Fe-N(heme) line (two conformations, named I and II, respectively); the two imidazole planes can be orthogonal to each other, parallel either to the two orthogonal C_{meso}-Fe-C_{meso} lines, or parallel to the two orthogonal N-Fe-N lines (two conformations, named III and IV, respectively); finally, the imidazole can have planes oriented at 45°, one parallel to the C_{meso}-Fe-C_{meso} line and the other parallel to the N-Fe-N line (one conformation, named V). In both the oxidation states, conformation II has been found to be the most stable and the other conformations have therefore been discarded for the successive calculations.

Finally, the vinyl, methyl and propionate side chains have been added in a standard geometry to the heme conformation II to construct a model of protoporphyrin IX-(imidazole)$_2$ complex. The geometry of this model has been optimized with a 6-31G* basis-set in both the oxidation states. The resulting molecular orbitals have been used for the subsequent RESP calculations.

The choice of equivalent atoms on which constraining identical point charges in the RESP calculation has been done is according to the AMBER force field (Cornell et al., 1995), as included in the NWCHEM package. All the atoms having the same point charge in the above database have been kept equivalent in the RESP calculations. The excess of charge obtained by linking the Cβ of His in the δ-protonated form to the imidazole has been equally spread over the Cβ and Hβ atoms of His-39 and His-63.

By comparing the results of RESP calculations for conformation II of reduced and oxidized forms, the major delocalization of the +1 charge excess is on Fe, on the four N heme atoms, and on the Hε1 and Hδ2 atoms of the imidazole ligands. Therefore, only those point charges localized on these atoms have been changed in the oxidized form with respect to the reduced form, adding +0.4 charge to Fe, +0.1 to each N heme atom, and +0.05 to each of the Hε1 and Hδ2 atoms of His-39 and His-63. The point charges of the other atoms of heme and His ligands have been kept identical to the reduced form. Point charges in reduced and oxidized forms are provided as Supplementary Material.

Molecular dynamics (MD) trajectories have been simulated for the reduced and oxidized forms of Cyt b_5. Both simulations started from the minimized averaged NMR structures (Arnesano et al., 1998a; Banci et al., 1997a) of each oxidation state (1AQA and 1AW3 PDB entries, respectively) with the missing starting and final residues added according to standard geometries. The heme moiety is taken in the A conformation with respect to the protein matrix in both the oxidation states. The net charges of residues have been chosen as they are at pH = 7, i.e., Glu, Asp, and the heme propionates carry a net −1 charge (in electron units of charge) and Lys and Arg +1. These starting structures have been energy-minimized in the vacuum. The force field has been taken from the AMBER database included in the NWCHEM package, except for the point charges modified according to the procedure summarized above. The stretching and bending parameters for the heme-His linkage have been taken identical to the C-N-Fe and N-Fe-N environments in the heme (Arnesano et al., 1998a; Banci et al., 1997b), whereas no energy restraint is applied on the orientation of the His imidazole planes with respect to the heme. The latter choice has been motivated by the need for introducing as little bias as possible in defining the heme orientation with respect to the protein matrix.

A number of 10 and 9 sodium ions have been added to the starting structures of reduced and oxidized forms, respectively, to balance the net negative charges on the two oxidation states. The ions have been located close to the most negative regions for the electrostatic potential computed on the molecular surface. The resulting structures have been merged into cubic boxes of TIP3P water molecules (Jorgensen et al., 1983). The side of the cubic boxes have been set to have a distance of 1 nm between the solute atom farthest from the solute center and the closest box side. All water molecules with the oxygen atom closer than 0.25 nm from any solute atom have been removed. An amount of 4371 and 3690 water molecules have been added to the reduced and oxidized forms, respectively, and the box sides were 5.25 and 4.97 nm.

The simulation boxes with the protein, sodium ions and water molecules have been energy-minimized. In the following MD simulations a thermostat and a barostat (Berendsen et al., 1984) have been used to keep the statistics close to the NTP ensemble corresponding to the given pressure and temperature. The pressure was always set to 0.1 MPa, and the compressibility kept isotropic to keep the simulation box cubic. The SHAKE algorithm (Ryckaert et al., 1977) was used to constrain the lengths of bonds involving hydrogen atoms to their equilibrium values, with a time step of 2 fs. A cutoff of 1 nm was used for the Lennard-Jones and direct-space part of electrostatic interactions, and the smooth particle-mesh Ewald algorithm (Essmann et al., 1995) was used for evaluating electrostatic interactions. The solvent molecules were initially simulated at the temperature of 300 K, then all systems were gradually heated from 0 to 300 K, with 5.8 ns of MD simulations at 300 K performed for both reduced and oxidized systems.

The analysis of the time evolution of the radius of gyration (data not shown) suggested that a rather long period of the trajectory is needed for the system to reach equilibration in the given force field. The first 1.3 ns of the trajectories have been excluded for the following analysis and the last 4.5 ns (45,000 configurations) have been kept. Moreover, we have computed all the following averages including 9000 configurations in the statistics, since the distribution of square radius of gyration and the first 10 first-rank relaxation modes (see below) did not change significantly reducing the number of configurations by a factor 5.

To analyze in detail the charge density around the iron ion, we make use of a suitable function emphasizing the changes of the radial part of the charge density with the oxidation state. Once the distributions $P_i(r)$ of distances between charged atoms and Fe have been computed for each kind i of charged atom or group, each distribution can be multiplied by the charge carried by the related atom or group and the resulting charge density can be properly normalized, as

$$g_c(r) = \frac{\sum_{i=1}^{M} P_i(r)\, z_i}{P_{id}(r)}, \tag{1}$$

where M is the number of charge types summarizing the charge density, z_i is the point charge carried by each type of charge, and $P_{id}(r)$ is the ideal distance distribution as it is obtained for a uniform density of charge carriers in the same shell at distance r from Fe.

We shall compute the above g_c function using both the full set of atomic point charges and a reduced set of point charges, assuming that the major contribution to the charge density is due to side chains carrying net charges and sodium ions. Four kinds of charges will be assumed to represent the entire charge density:

1. The negative charge of carboxylate groups, located on the two carboxylate oxygens of Glu and Asp residues ($z_1 = -1/2$) and on the heme propionate oxygen atoms (46 atoms).
2. The positive charge of guanidinic groups of Arg, located on the two Nη atoms ($z_2 = 1/2$) (6 atoms).
3. The positive charge of ammonium groups of Lys, located on the Nζ atom ($z_3 = 1$) (10 atoms).
4. The charge of the sodium ions, which carry the charge $z_4 = 1$ (10 and 9 atoms in the reduced and oxidized forms, respectively).

The distance distributions calculated for distance pairs involving Fe and these groups have been averaged over all the equivalent groups in the protein and divided by the same distribution of the corresponding ions as uniformly distributed in the average volume of the simulation box.

Diffusion theory and diffusive models

Nuclear relaxation rates of Cyt b_5 in both the oxidation states were experimentally determined to understand the extent of configurational fluctuations in the two redox states (Banci et al., 1998a; Dangi et al., 1998a). Particularly relevant NMR relaxation parameters are the relaxation rates of the z-component of the magnetization due to ^{15}N nuclei in the static magnetic field ($R(N_z)$, or R_1, hereafter) and in the rotating frame ($R(N_{z'})$, or $R_{1\rho}$ hereafter). The $R_{1\rho}$ parameter is more sensitive to slow motions affecting the protein backbone: its frequency dependence can be experimentally determined, when affected by characteristic times in the range of milli- and microseconds, due to the setup of the NMR experiments.

In this work we shall ignore the effects of paramagnetic atoms on the above relaxation rates. These effects have been estimated to be at maximum 0.14 and 0.17 Hz for $R(N_z)$ and $R(N_{xy})$ (or R_2), respectively, over experimental average values of ~2 and 9 Hz for R_1 and $R_{1\rho}$, respectively, in the oxidized form (Banci et al., 1998a; Dangi et al., 1998a). Therefore, the paramagnetic effect accounts for a maximum variation of <10% in the measured data of the oxidized form.

The relationships between the above NMR relaxation rates and spectral densities are given by the following equations (Peng and Wagner, 1992):

$$R_1 = dJ(\omega_H - \omega_N) + (3d + c)J(\omega_N) + 6dJ(\omega_H + \omega_N); \tag{2}$$

$$R_{1\rho} = d\{4\sin^2(\beta)J(\omega_e) + [\sin^4(\frac{\beta}{2})J(\omega_H - \omega_N + \omega_e) + \cos^4(\frac{\beta}{2})J(\omega_H - \omega_N - \omega_e)] + 3[\sin^4(\frac{\beta}{2})J(\omega_N - \omega_e) + \cos^4(\frac{\beta}{2})J(\omega_N + \omega_e)] + \frac{3}{2}\sin^2(\beta)[J(\omega_N + \omega_e) + J(\omega_N - \omega_e)] + 6\cos^4(\frac{\beta}{2})J(\omega_H + \omega_N + \omega_e)\} + c\{\frac{2}{3}\sin^2(\beta)J(\omega_e) + [\sin^4(\frac{\beta}{2})J(\omega_N - \omega_e) + \cos^4(\frac{\beta}{2})J(\omega_N + \omega_e)]\}. \quad (3)]$$

The meaning of the symbols contained in the above equations is

$$d = \frac{\gamma_N^2\gamma_H^2\hbar^2}{20}\langle 1/r^6\rangle, \quad (4)$$

where γ_i is the gyromagnetic ratio of nucleus i and r is the modulus of the given N-H vector that, in our model, has been kept constant (see previous subsection),

$$c = \frac{(\delta\,\omega_N)^2}{15}, \quad (5)$$

where δ is the chemical shift anisotropy of each nucleus N,

$$\beta = \arctan(\frac{\omega_1}{\Delta}), \quad (6)$$

where ω_1 is half of the radio-frequency amplitude in the plane perpendicular to the static magnetic field and Δ is the chemical shift in rad/s (Larmor frequency) of each N nucleus, and finally,

$$\omega_e = \sqrt{\omega_1^2 + \Delta^2}. \quad (7)$$

In the above equations, we have assumed that ^{15}N nuclear spin relaxation is governed by the modulation of the dipolar coupling with its bonded ^{1}H and of the chemical shift anisotropy. The time dependence of both these interactions is determined by the same process: the spectral densities J in the equations above contain the stochastic motions of the N-H bond, the movement of this vector also being assumed to describe the motions of the principal axis of the chemical shift tensor. We have assumed $\delta = 160$ ppm for all the N nuclei in the protein. The contribution of the chemical shift exchange or modulation (Desvaux et al., 1995) can be, in principle, included. It is here neglected because of the need of computationally expensive models for including the functional dependence of ^{15}N chemical shift from the atomic configuration (Xu and Case, 2002).

The relationship between J and the time correlation function (*TCF*, hereafter) statistically describing the contribution of the molecular dynamics on spin transitions is given by Cavanagh et al. (1996), as

$$J(\omega) = 2\int_0^\infty \cos(\omega t)\,TCF(t)dt. \quad (8)$$

Finally, the TCF is given by

$$TCF(t) = \sum_{M=-2}^{2}\langle[D^{(2)*}_{M,0}(\Omega(t))][D^{(2)}_{M,0}(\Omega(0))]\rangle = \langle P_2(\cos(\beta(t)))\rangle, \quad (9)$$

where $D^{(2)}_{M,0}$ are irreducible spherical tensors (Rose, 1957) and Ω is the direction of the given N-H vector. P_2 is the Legendre polynomial of order 2, and β is here the angle that the N-H vector spans in time t.

A quantity that summarizes the mobility of the N-H vector is the integral of the orientational part of the TCF,

$$\tau = \int_0^\infty TCF(t)dt. \quad (10)$$

The quantity τ is called correlation time of the given internuclear unit vector.

The mode-coupling diffusion (MCD) theory of the dynamics of a biological macromolecule in solution is adopted here for the computation of TCFs. The diffusion theory treats the solvent hydrodynamically and uses a detailed molecular model for the macromolecule in terms of beads (atoms or groups of atoms) connected by real or effective bonds diffusing in an interatomic potential (the same used in the MD simulation). The beads are represented as points of coordinates $\mathbf{r}_i$, with i running over the N_a beads in the frictional model, and friction coefficients $\zeta_i = 6\pi\eta\,a_i$, with η as the solvent viscosity. The Stokes' radii a_i of the beads are calculated here by using the solvent-accessible surface area (SASA) method (Pastor and Karplus, 1988) with a zero probe radius, by summing the surfaces of each constituent group (La Penna et al., 2000b).

In this work, we shall use two kinds of bead representations for the macromolecule. The macromolecular model with beads located on solute atoms (the ''bare'' model, hereafter) uses one bead for each amino acid (positioned on the Cα atom) and one bead for the heme (positioned on Fe; La Penna et al., 2000a). In the second model, the beads are positioned on the water molecules surrounding the macromolecule (the ''surface'' model, hereafter) and all the beads have the same Stokes' radius of 0.14 nm (i.e., the usual Stokes' radius of the water molecule).

These bead frictional models, both with and without the inclusion of water layers, have been widely used for rigid macromolecules (de la Torre et al., 2000; Fernandes et al., 2002). The MCD approach in the form summarized below constitutes the generalization of diffusion theory to flexible macromolecules that gives the exact rotational diffusion in the limit of rigid macromolecules (La Penna et al., 1999).

The MCD approach (La Penna et al., 1999; Perico and Pratolongo, 1997), can be briefly summarized as follows. The N_a beads of the given frictional model can be connected by N_b bonds ($\mathbf{l}_i$, $i = 1, \ldots, N_b$). The variables $\mathbf{l}_i$ can be organized in the $3 \times N_b$ dimensional array $\mathbf{l}$ containing all the bond vectors $\mathbf{l}_i$. This array entirely describes the model configuration, and its dynamics is regulated by the operator L, adjoint to the diffusion Smoluchowski operator D,

$$\frac{\partial\mathbf{l}}{\partial t} = L\mathbf{l}; \quad L = \sum_{i,j=1}^{N_a}[\nabla_i\mathbf{D}_{i,j}\nabla_j - (\nabla_i U/k_BT)\mathbf{D}_{i,j}\nabla_j], \quad (11)$$

where U is the potential energy of the beads as a function of the bead coordinates, k_B is the Boltzmann constant, and T is the absolute temperature.

The diffusion tensor $\mathbf{D}$ is given by

$$\mathbf{D}_{i,j} = D_i\,\mathbf{H}_{i,j}, \quad (12)$$

$$\mathbf{H}_{i,j} = \mathbf{1}\,\delta_{I,j} + \zeta_i\,\mathbf{T}_{i,j}(1 - \delta_{i,j}), \quad (13)$$

with **H** and **T** representing the hydrodynamic interaction matrix and the Rotne-Prager (Rotne and Prager, 1969) tensor, respectively, and $D_i = k_B T/\zeta_i$ is the diffusion coefficient of each bead.

By expanding the conditional probability (solution to the Smoluchowski equation) in a complete set of eigenfunctions of L, the TCF of any coordinate-dependent dynamic variable with zero average $f(t)$ may be expressed in the standard form

$$\langle f(t)f(0)\rangle = \sum_i \langle f\,\psi_i\rangle\langle\psi_i\,f\rangle \exp(-\lambda_i t), \qquad (14)$$

where $-\lambda_i$ and ψ_i are, respectively, the eigenvalues and the normalized eigenfunctions of the operator L,

$$L\psi_i = -\lambda_i\psi_i. \qquad (15)$$

This eigenvalue equation is solved by expansion of the eigenfunctions in a suitable basis-set. Once a basis-set has been chosen, the eigenvalue equation for the diffusion operator becomes a generalized eigenvalue equation in matrix form that can be solved with standard methods. By using the RM2-II basis-set (i.e., a second-order approximation) in MCD theory (La Penna et al., 1999), the second-rank TCFs are sum of exponential functions, with the constant λ_i the rates of the relaxation modes. Note that the isotropic reorientational eigenmode dynamics method, mentioned in the previous section, gives an alternative derivation of these relaxation rates: the time correlation functions for the projection on the static eigenvectors of the covariance matrix of the second-rank tensor components related to relaxing vectors (f in Eq. 14) are derived directly from the MD simulation.

Once the second-rank TCFs have been computed with the RM2-II basis-set, they can be analytically Fourier-transformed to the spectral densities (Eq. 8) and analytically integrated to give correlation times (Eq. 10). In the MCD approach, the incomplete conformational sampling of MD affects the rates and the coefficients in the multiexponential expansion of the TCFs of interest (Eq. 14), introducing errors that increase with the rate.

RESULTS

Configurational statistics

The Fe atom in the heme group has been covalently bonded to the Nϵ2 atoms of His-39 and His-63 to mimic the Fe axial coordination in Cyt b_5, but no torsional potential contributions related to all dihedral angles centered on the Fe-Nϵ2 bonds have been included. Therefore, it is meaningful to analyze the deviations of these dihedral angles from the NMR structures both in the reduced and oxidized states. In the reduced form, the dihedral angles NC-Fe-Nϵ2-Cϵ1 in His-39 and His-63 are -43 ± 9 and $32 \pm 13°$, respectively, averaged over the MD trajectory, which compare with –45 and 25° in the most refined available solution structure (1B5A PDB entry) (Dangi et al., 1998a).

For the oxidized form, the averages over the MD trajectory of the same dihedral angles are -33 ± 10 and $33 \pm 11°$, respectively, compared to -39 and 21° of the available crystallographic structure (1CYO PDB entry) (Durley and Mathews, 1996) and -45 and 24° of the NMR-minimized average structure (1AW3 PDB entry) (Arnesano et al., 1998a). For both the redox forms, only moderately larger average dihedral angles are observed compared to NMR data. On average, the MD trajectories of both the redox forms display the two imidazole planes parallel to each other and only slightly rotated from the CB-Fe-CD *meso* direction (conformation II in Methods, corresponding to -45 and $+45°$) toward the NC-Fe-NA direction. Conformation II has been found the most stable by ab initio calculations (see Methods) and both MD-averaged dihedral angles are consistent with this conformation within statistical errors. Moreover, in the oxidized form the MD-averaged dihedral angles are consistent with the values derived by paramagnetic effects on NMR parameters (see Fig. 5 in Banci et al., 2002).

Among the specific noncovalent interactions between the heme group and the protein, the most important difference between the two redox forms is in the behavior of the heme propionate A. In the reduced form, the carboxylate group of propionate A is highly screened by a sodium counterion and it does not interact specifically with any protein atom (data not shown). On the other hand, in the oxidized form the carboxylate oxygen atoms are involved in strong hydrogen bonds with the backbone amidic H atoms of Ser-64 and His-63. The average N(Ser-64)-OA distance is 0.35 ± 0.07 nm in the oxidized form, compared to 0.7 ± 0.1 nm in the reduced form. The effect of these hydrogen bonds is the bending of the propionate side chain toward Fe and the heme center in the oxidized state. The differences in the conformation of propionate A are fully consistent with the experimental solution structures (Arnesano et al., 1998a; Banci et al., 1997a; Dangi et al., 1998b).

The effect of the different oxidation state on the secondary structure can be monitored through the time evolution of the state of each residue along the MD trajectories (Kabsch and Sander, 1983). In Fig. 1, every 45 ps for each residue the helical state is identified by a solid bar, with the β-sheet state by a shaded bar, and compared to α-helices in the NMR structures, shown as vertical bars on the left. It can be observed that h1, h3, h4, and h5 are all maintained during the simulation, showing a larger stability of h3 and h4 in the oxidized state with respect to the reduced state. The region corresponding to h2, a short helical portion in both NMR structures, is disordered after MD equilibration, whereas h6 is partially lost in the oxidized state only. In both redox forms, the β-sheet structure is well-maintained during the whole MD simulation. Except for h2 and h6, the α-helical and β-sheet populations obtained by the MD trajectories represent the wobbling of the secondary structure elements in both reduced and oxidized states solution structures.

The disruption of h2 (region 33–38) is also observed by partial unfolding of the oxidized form of Cyt b_5 in 2 M guanidinium chloride (GdmCl) (Arnesano et al., 1998b, 2000), thus confirming the low stability of this protein region. The behavior of the second ''breaking point,'' region 62–64 located in the loop between h4 and h5, is also expected to be affected by GdmCl: the strong hydrogen bonds of Ser-64 and His-63 with the heme propionate A (see above) are weakened by increasing GdmCl concentration,

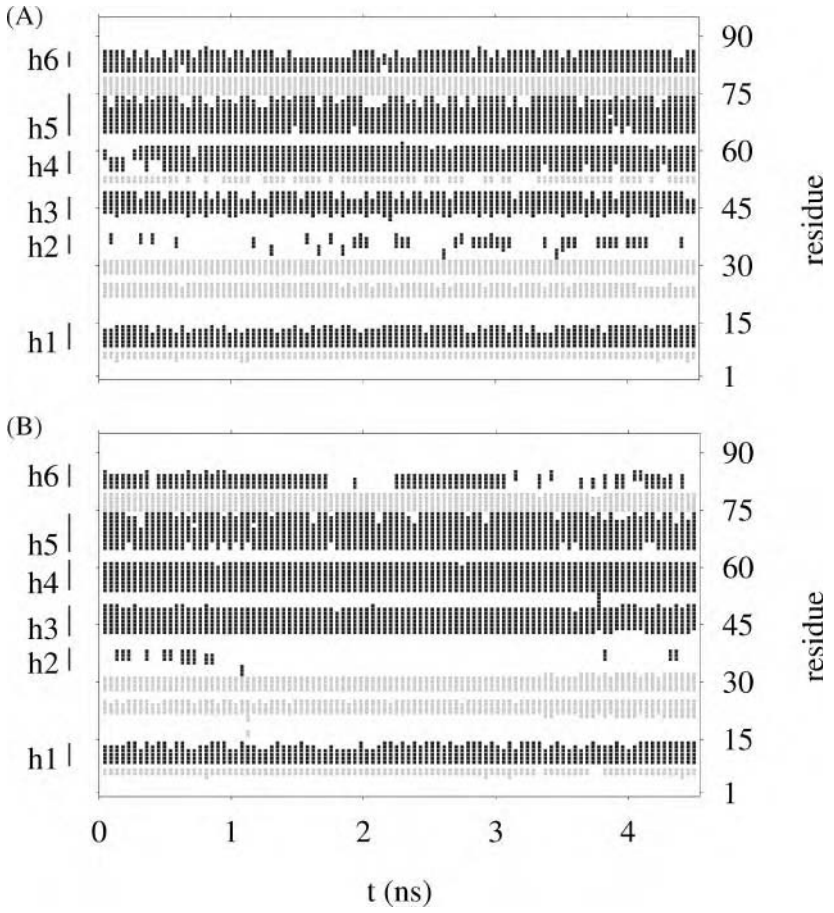


FIGURE 1 Time evolution of secondary structure of reduced (*A*) and oxidized (*B*) forms along the MD simulations: helices are identified in solid bar, β-sheets in shaded bar; all the other motives are not displayed. Time resolution is 45 ps. Residues in α-helix in the NMR structures are displayed as vertical bars on the left.

thus partially unfolding the native structure with slight influence on the secondary motives in the nearby helices (h4 and h5).

Because of the slow decay through space of electrostatic interactions, the modification of charge density on the heme group upon oxidation is expected to produce long-range effects both on the protein matrix and on the solvent and ions surrounding the protein. This may be particularly important in Cyt b_5 because the heme is close to flexible portions of the protein and to the bulk polar environment (water and counterions), where charged atoms are allowed to move upon change in the electric field. The residues more sensitive to a change of charge density in the heme are those containing a net charge in the side chains, i.e., Glu–, Asp–, Lys+, and Arg+. Any structural change affecting these charged residues is also expected to affect, more or less directly, the solvent and ionic protein environment. This long-range effect can significantly contribute to the variation of entropy upon heme oxidation.

In Cyt b_5 at pH 7 there are 21 negatively charged (9 Asp and 12 Glu) and 13 positively charged (10 Lys and 3 Arg) residues. The further addition of the -2 and -1 charges of the heme in the reduced and oxidized forms, respectively, produces a net negative charge of -10 and -9 in the two oxidation states, respectively. In our model the negative charge in the molecule has been balanced by 10 and 9 sodium ions in the reduced and oxidized forms, respectively. To summarize the different behavior of the negative and positive charge locations in the two redox forms of the protein, we have computed a set of charge density distribution functions for the full atomic point charges used in the MD simulation and for simplified representations of net charges.

Assuming that the heme group did not significantly change its position within its pocket, as it is revealed by the very minor changes in the geometrical parameters with respect to the His ligands, we analyzed the radial part of the charge density distribution (i.e., $g_c(r)$ of Eq. 1) in terms of distances between Fe and any other charged point in the protein and sodium counterions.

The $g_c(r)$ function is plotted in Fig. 2 for the two redox forms of the protein and using the atomic point charges. It can be noticed that the function displays a rugged behavior in the range 0–1 nm, which is a direct effect of the charge density modification within the heme pocket. The peaks within 0.5 and 1 nm contain the contributions of the heme propionates (see below) and the smoother behavior of the charge density beyond 1 nm is due to residues in the protein

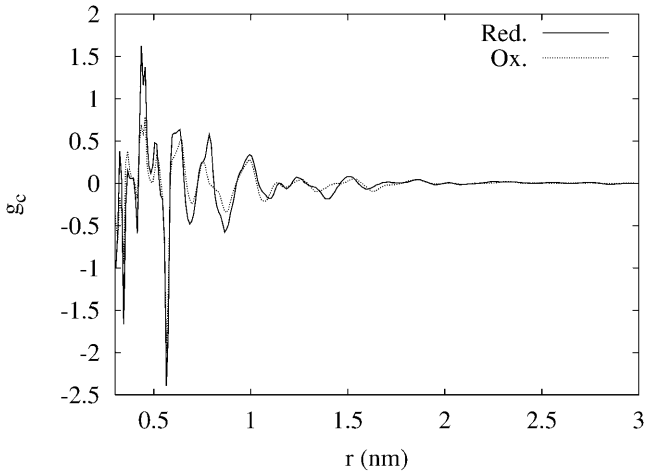


FIGURE 2 Charge radial distribution function of Eq. 1 in the text, calculated with the atomic point charges. Reduced form (*solid line*) and oxidized form (*dotted line*).

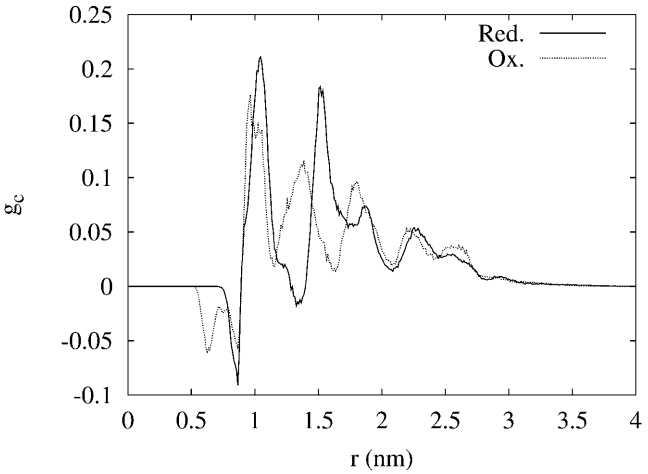


FIGURE 3 Total charge radial distribution functions calculated with simplified point charges (see text). Reduced form (*solid line*) and oxidized form (*dotted line*).

matrix not directly bonded to the heme. Therefore, in this latter region the changes in charge density with oxidation is due to the structural change of the protein matrix and of the counterion distribution around the heme pocket. It can be observed that a significant change in charge density occurs at ~1.5 nm from Fe, where a negative contribution in the reduced form is compensated by some positive contribution in the oxidized form.

It must be noticed that the heme group in Cyt b_5 is almost on the protein-water interface with the charged propionates on the entrance of the heme pocket. This means that most of the charges represented in the $g_c(r)$ function are on the opposite side with respect to the heme-solvent interface.

To separate the various contributions to charge density, we have calculated the charge density contributions of those classes of groups expected to be more relevant for understanding the charge density modifications, i.e., the groups of residues carrying a net negative or positive charge and the sodium counterions (see Methods). In Fig. 3, the $g_c(r)$ function for this reduced charge-set is plotted for the two redox forms of Cyt b_5. Different effects of the redox state of Fe can be observed in the three r ranges 0–1 nm, 1–2 nm, and beyond 2 nm. In the first range the closest negative density in the oxidized form is due to the bending of one of the two heme propionates. In the reduced form one of the 10 sodium ions makes a strong salt bridge with the A propionate (data not shown), but no sodium ion can come too close to Fe in the oxidized form because of the strong hydrogen bonds between the propionate and surrounding atoms, i.e., the backbone amidic H atoms of Ser-64 and His-63. In the second range, $1 < r < 2$ nm, the positive charge density significantly changes upon heme oxidation beyond 1.2 nm. In this region, a decomposition of g_c in terms of positive and negative contributions shows that negative charges move farther from Fe and positive charges become closer to Fe in the oxidized state, i.e., the opposite of what is expected. The charge movement inverts the density oscillation in the 1–2 nm region and this behavior accounts for the change in charge density in the model observed at distances ~1.5 nm (Fig. 2). Beyond 2 nm, the effect of the change in oxidation state becomes almost negligible.

A possible explanation of the charge density change occurring with oxidation can now be given. In the reduced form the negative charge closest to the heme group, i.e., the propionates, are partially screened by positive mobile counterions (in our model the sodium ions) that are allowed to come close to the heme because of the structure of the heme pocket and the less positive charge on the heme itself. This screening allows other negative charges to come close to the heme group. With oxidation of the heme and with the consequent increase of positive charge on it, the propionates slightly move toward the center of the heme and the heme pocket is narrowed by other interactions like those between propionate A and the amidic H of His-63 and Ser-64, the latter being particularly strong. The sodium ions cannot come close to the heme and the negative charges of the propionates become less screened. In the oxidized form, the negative charges close to the heme more efficiently repel negative charges at larger distances (~1.5 nm) pushing negative charges in this region farther from the heme, whereas they attract positive charges at about the same distance.

The molecular surface can be indirectly changed by the electrostatic interactions because of possible change in the distribution of hydrophobic and hydrophilic groups on the surface accompanying the charge density modification. This effect can be observed by looking at the time evolution of the SASA computed for the whole molecule in the two different oxidation states (Fig. 4) (Eisenhaber et al., 1995). It can be observed that the molecular surface in the oxidized state

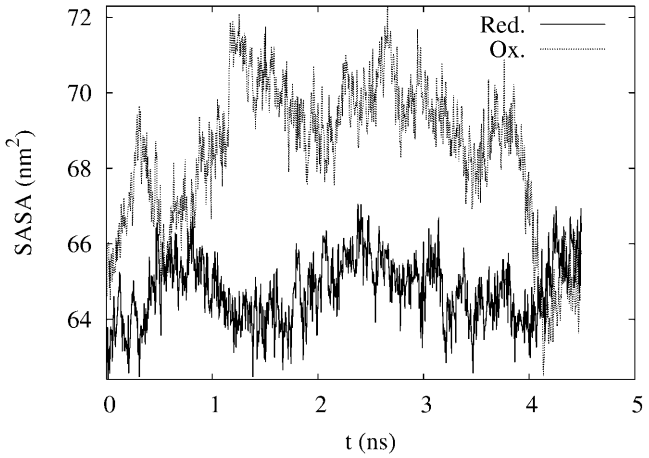


FIGURE 4 Time evolution of molecular SASA for the reduced (*solid line*) and oxidized (*dotted line*) forms.

tends to increase during the MD trajectory. Starting from a value very similar to that of the reduced form, it becomes ~10% larger after 1 ns and back to the reduced form values at the end of the trajectory. This behavior shows that the molecule in the oxidized form is changing shape during the trajectory whereas in the reduced state the starting shape is more stable. A comparison with the time evolution of the secondary structure (see Fig. 1) shows that the molecular expansion in the oxidized form occurs when h2 is not populated.

The two sets of observations above—the first concerning the statistical change of charge density and the second concerning the change in shape of the protein matrix, occurring in the oxidized form with respect to the reduced form—can be now linked together. The molecular expansion is mainly related to a conformational change in the C-terminal region. In Fig. 5, the configurations of reduced and oxidized forms are represented at the time where the molecular SASA increase is approximately at its maximum ($t = 3$ ns). The different position of the C-terminal residues 83–94, containing part of h6, in the oxidized form is evident: the distance between this region and the main protein body increases significantly. In the oxidized form, this configuration displays the side chains of the negative residues Asp-53, Glu-56, and Glu-59 farther from the side chains of residues Arg-84, Lys-86, and Lys-89 compared to the same configuration in the reduced form (data not shown). The breaking of electrostatic interactions between these two sets of opposite charges is related to the structural change involving the residues 83–94 in the oxidized state. Moreover, this event exposes the high density of negatively charged residues in the region 43–60 to the solvent. The latter region has been identified as one of the negative patches of Cyt b_5 involved in the interaction with the positive patch of Cyt c (Banci et al., 2004; Qian et al., 2001; Rodgers et al., 1988; Wu et al., 2001). Finally, significant deviations from the x-ray structure have been observed in this region in MD simulations of the bovine liver Cyt b_5 (Storch and Daggett, 1995), suggesting the importance of a modulation in solvent accessibility of the negative patch for Cyt b_5-protein interaction. Indeed, in the bovine liver oxidized Cyt b_5 crystallographic structure (Durley and Mathews, 1996), residues 89–94 were not visible in the electron density map and were thought to have multiple conformations. Therefore, the movement of the C-terminal region is an indirect long-range and global effect of the change in charge density due to the change in oxidation state of iron that can modulate the Cyt b_5 affinity for Cyt c.

All the structural modifications described above for the oxidized form are within 0.3 nm (estimated over Cα of residues 1–88) from the x-ray bovine structure (Durley and Mathews, 1996) and from rat microsomal NMR structure (Arnesano et al., 1998a). This range of variability is commonly observed as a consequence of the actual solution environment at room conditions.

To conclude the analysis of conformational changes occurring in Cyt b_5 upon oxidation, we computed the solute-solvent radial distribution function (RDF, hereafter), $g(r)$, following a reported procedure (La Penna et al., 2003). This function contains relevant information on the structure of the water shells surrounding the protein. In Fig. 6, the RDF function of distance pairs involving any of the protein atoms and any of the water oxygen atoms have been plotted

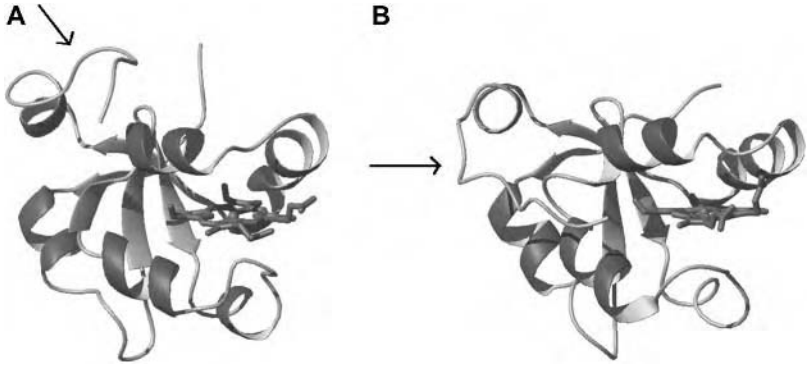


FIGURE 5 Configurations of reduced (*left*) and oxidized (*right*) forms at 3 ns, corresponding to about the maximum increase of molecular SASA in the oxidized form. The heme is displayed with all the bonds connecting heavy atoms and Fe is displayed as a sphere. Arrows point to the C-terminus. Figure prepared with the MolMol program (Koradi et al., 1996).

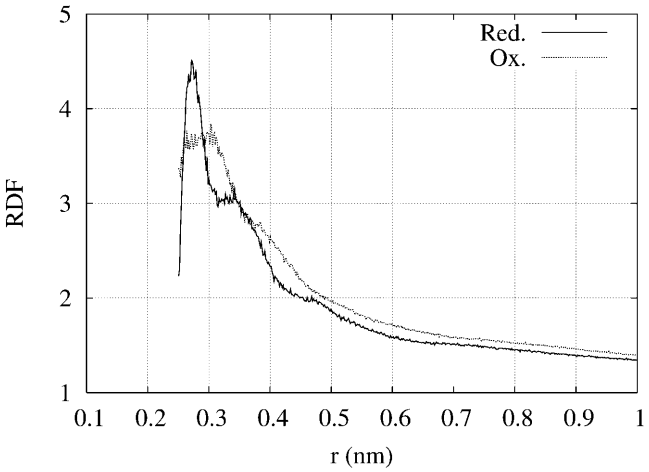


FIGURE 6 Solute-solvent radial distribution function (RDF) for the reduced (*solid line*) and oxidized (*dotted line*) forms.

for the reduced and oxidized forms. As already observed (La Penna et al., 2003), the first peak of the function contains pairs involving heavy protein atoms and the water oxygen in the first solvent shell. In both cases, for large r-values the RDF approaches values >1 because of the higher density of water as obtained in the simulation box with respect to bulk water at the same temperature and pressure conditions. The most significant difference in the RDF between the two redox forms is, however, concentrated in the first shell: in the reduced form, a well-structured water shell can be observed with water molecules ordered within 0.3 nm; in the oxidized form, this water shell is looser, spreading up to 0.35 nm far from the solute, and the distinction is no longer possible between the first and second solvation shells. The less-defined first water shell in the oxidized form is related to the more fluctuating molecular surface. As observed for other proteins and nucleic acids, the radial organization of the first water shell is mainly due to hydrophilic and charged groups exposed to the solvent by the molecular surface. In the oxidized form of Cyt b_5, these groups are more mobile, as is shown by the wider distribution of both the Fe-negative and Fe-positive radial charge distributions displayed in Fig. 3. Therefore, a water shell has fewer chances to be created in the oxidized form. All these observations suggest that the decrease in water organization, as it occurs in the MD simulation in explicit solvent, can be relevant for explaining the positive entropy change upon oxidation of Cyt b_5 (Dangi et al., 1998a). The implications of these observations can be consistently included in the following diffusive model.

Molecular dynamics and interpretation of NMR experiments

The information acquired on the configurational statistics and summarized in the previous subsection can be included in the calculation of nuclear spin relaxation parameters by using diffusion theory and diffusive models based on the different statistics of friction points. As explained in Methods, two hydrodynamic models will be applied here. In the bare model each residue is modeled as a unique friction point with a Stokes' radius derived by the 0-probe SASA of the entire residue. The friction points are located on the Cα of each residue and on Fe for the heme group. In the surface model, friction points are located on the molecular surface as it is accessed by the solvent, i.e., on the surface with area measured by the molecular SASA (see above). The choice of the second model is motivated by the observed modification of the molecular surface properties with oxidation and by the importance of the structural changes in the first shell of water surrounding the protein. The density of friction points in this latter model will be set to the density of water molecules as it is calculated directly from the MD trajectory in the explicit solvent. The density of water molecules in the first solvation shell is the integral of the distribution used in computing the solute-solvent RDF in Fig. 6 in the range where the first maximum in the RDF is contained.

In the reduced form, the first peak in RDF is within 0.30 nm, and 360 $\pm$ 13 water molecules have been counted on average in this range. In the oxidized form, the first peak is approximately in the range 0–0.35 nm, where 530 $\pm$ 14 molecules are located. Therefore, the surface model for the reduced form is made of 360 friction points located on the surface, whereas in the oxidized form it is made of 530 friction points. Once the water density has been explicitly calculated from the MD trajectory, friction locations are chosen randomly on the molecular SASA, as already described in La Penna et al. (2003).

By analyzing the first-rank rates obtained by the diffusive bare and surface models, it is possible to derive estimates of the diffusion tensor principal components of a molecule with an average shape derived by the MD trajectory. The protein in both redox states has $\lambda_1 < \lambda_2 \sim \lambda_3$, thus representing a diffusion tensor with eigenvalues in the order $D_{zz} > D_{xx} \sim D_{yy}$ and a rod-like diffusion tensor. By using the relationships in Table 2 of La Penna et al. (1999), we obtain the results in Table 1 (this article). The ratio $D_{\|}/D_{\perp}$ for both models is comparable with the estimates in the literature (Dangi et al., 1998a), even if a slightly less anisotropic diffusion tensor is found here for the oxidized form (1.30 compared to 1.35). As expected, the diffusion tensor eigenvalues decrease by using the solvated diffusive models compared to the bare models, because of the larger size of the

TABLE 1 Diffusion tensor eigenvalues ($\times 10^{-7}$ s^{-1}) estimated from diffusion rates and their ratio

Model	Reduced bare	Oxidized bare	Reduced surface	Oxidized surface
$D_{\|}$	4.56	4.69	2.69	2.54
$D_{\perp}$	3.66	3.60	2.17	2.05
$D_{\|}/D_{\perp}$	1.26	1.30	1.24	1.24

surface bead aggregates. However, the ratio does not change with the model and the rotational diffusion anisotropies of the two redox forms are not significantly different.

To analyze the behavior of the mobility of the N-H bonds governing the ^{15}N relaxation rates, we first computed the integral of the $P_2(t)$ TCFs for each of the backbone amidic bonds in the protein (i.e., the correlation times τ of Eq. 10). In Fig. 7, the correlation times have been plotted for the two redox forms calculated with the ''bare'' model. For both the redox forms, the correlation times are smaller in the turns and loops connecting the helices than in the helices themselves, with the exception of h2 (residues 33–38), which is rarely populated in the MD simulations. The correlation times in the oxidized forms are significantly smaller than in the reduced form in at least four protein regions: residues 18–19, h2, the first residues in h3 and in the C-terminal region, starting from the end of h6 (residues 86–94). On the other hand, His^{63} and the residues in the nearby sequence are more rigid in the oxidized form than in the reduced form, because of the strong interaction with the heme A propionate.

Summarizing, the mobility obtained by these correlation times is consistent with other structural distributions observed in the previous subsection. The α-helical populations of residues in the two redox forms show that h2 population is almost negligible and lower in the oxidized than in the reduced form. Similarly, the higher mobility of the C-terminus in the oxidized form reveals the wide range of configurations achieved by this region in the oxidized state (see above).

A direct comparison of R_1 ^{15}N relaxivities can be made between the values calculated through the two diffusive models and experiments. The R_1 values for reduced and oxidized forms are shown in Fig. 8, *A* and *B*, respectively. It can be observed that the choice of the diffusive model produces mainly a shift of the data because of the global change of relaxation rates. In both redox states the experimental data show a smaller variation along the protein chain compared to the calculated values, particularly in the regions that have been found less ordered in the simulations. The diffusive method allows the calculation of the $P_2(t)$ TCF as the sum of many exponential functions, each decaying with one of the eigenvalues (rates) of Langevin equation in its matrix representation. The simplification of such TCF in terms of two exponential functions is possible and the amplitude of the function with the lowest rate is the so-called Lipari-Szabo order parameter (Fausti et al., 1999). For this protein, this simplification gives almost the same NMR relaxivities of using the entire set of rates (data not shown), thus showing that for these molecular statistics the information on the mobility pattern is all contained in the order parameters and in the five lowest second-rank relaxation modes. For both the models, the differences between experiments and calculations arise from the low convergence of the MD trajectories.

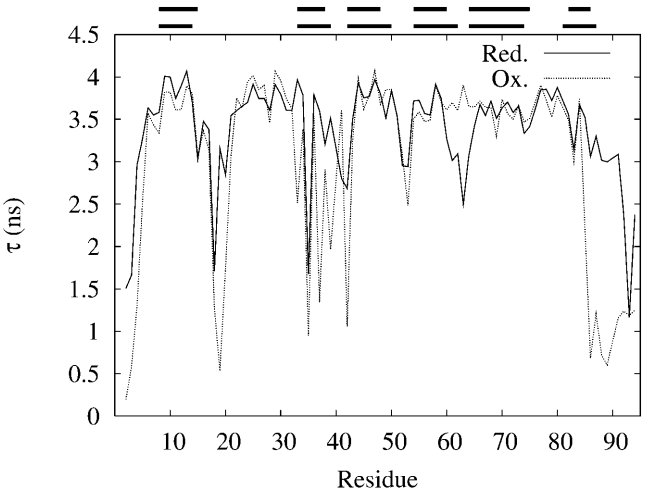


FIGURE 7 Correlation times (Eq. 10 in text) of H-N bonds in the reduced (*solid line*) and oxidized (*dotted line*) forms calculated with the bare model. Residues in α-helix in the NMR structures are displayed as horizontal bars for reduced (*top*) and oxidized (*bottom*) forms.

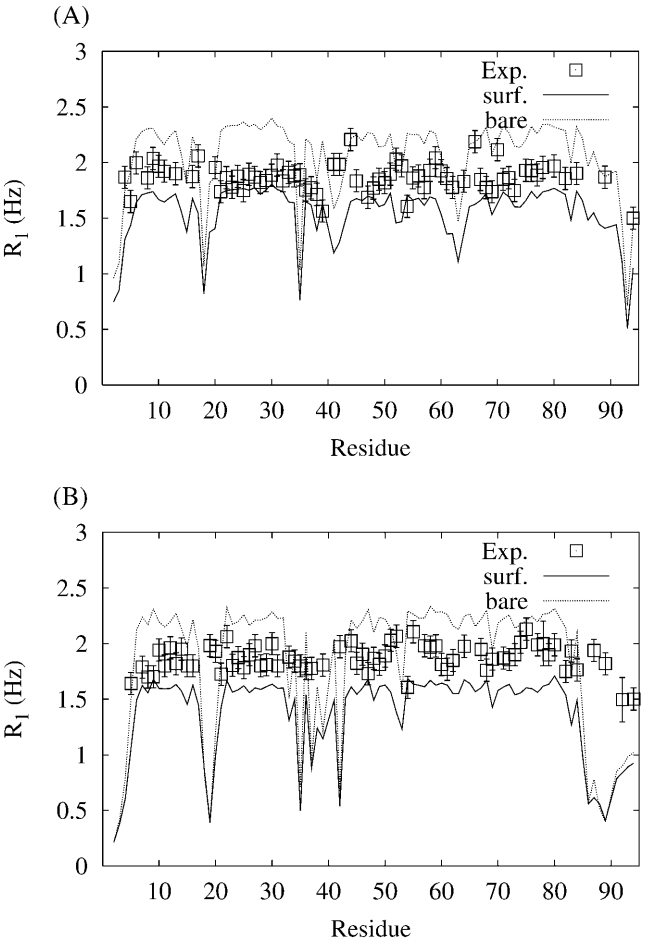


FIGURE 8 R_1 NMR relaxation rates of ^{15}N at $\nu(^1\text{H}) = 600$ MHz for the reduced (*A*) and oxidized (*B*) forms: experiments (*squares* with *error bars*), results of the bare model (*solid line*), and results of the surface model (*dotted line*).

A careful analysis of the order parameters shows that in the less ordered regions, like h2, the number of conformational transitions affecting the N-H bond orientation during the entire MD trajectories are only a few in number. The problem of the statistical accuracy of order parameters in these conditions has been discussed in the literature (Chandrasekhar et al., 1992). For systems undertaking rare configurational transitions, a more meaningful estimate of order parameters can be performed by averaging data using blocks of configurations not containing the transitions. In the present case, it has been observed that this condition is fulfilled by computing order parameters as averages of 15 independent time windows of data averaged over 300 ps each.

In Fig. 9, the order parameters obtained with and without block-averaging have been plotted for the two redox forms and compared with the order parameters obtained by fitting R_1 relaxation rates (Dangi et al., 1998a). It can be observed that rarely sampled conformational transitions in the less ordered regions dramatically affect the order parameters and, therefore, the mobility and the relaxation rates, but the systematic removal of these statistical errors can partially recover the experimental information and flatten the R_1 variation within the molecular chain.

In Fig. 10, the computed $R_{1\rho}$ data at $\omega_e = 1870$ Hz are compared with the experimental ones published in the literature (Banci et al., 1998a). As expected, the factor of two in the diffusion constants (see Table 1) is now appearing in the relaxation rates, being that $R_{1\rho}$ is more sensitive to the low frequency contributions to the spectral densities. In both the reduced and oxidized forms, the low $R_{1\rho}$ relaxation rates are captured by the bare model, whereas the high values are better reproduced with the surface model. The behavior of these relaxation rates with the position of the N atom in the chain is almost the same in the two models and it displays the same pattern shown by R_1. Both calculated patterns are governed by the order parameters of the N-H bond in the different redox forms of the protein.

Apart from the correct reproduction of the N-H bond mobility as it is contained in NMR experiments, it is interesting to notice that in Figs. 8 and 10 experimental data

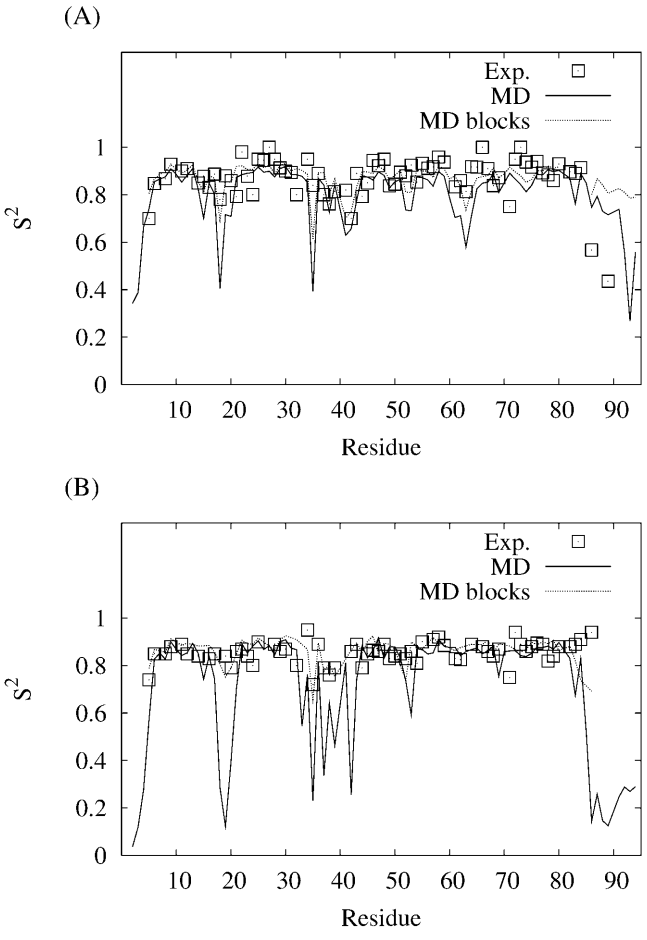


FIGURE 9 Lipari-Szabo order parameters for the reduced (*A*) and oxidized (*B*) forms: results obtained by fitting R_1 NMR experiments (*squares*), results from averaging the MD trajectory (*solid line*), and from averages with blocks of 300 ps each (*dotted line*).

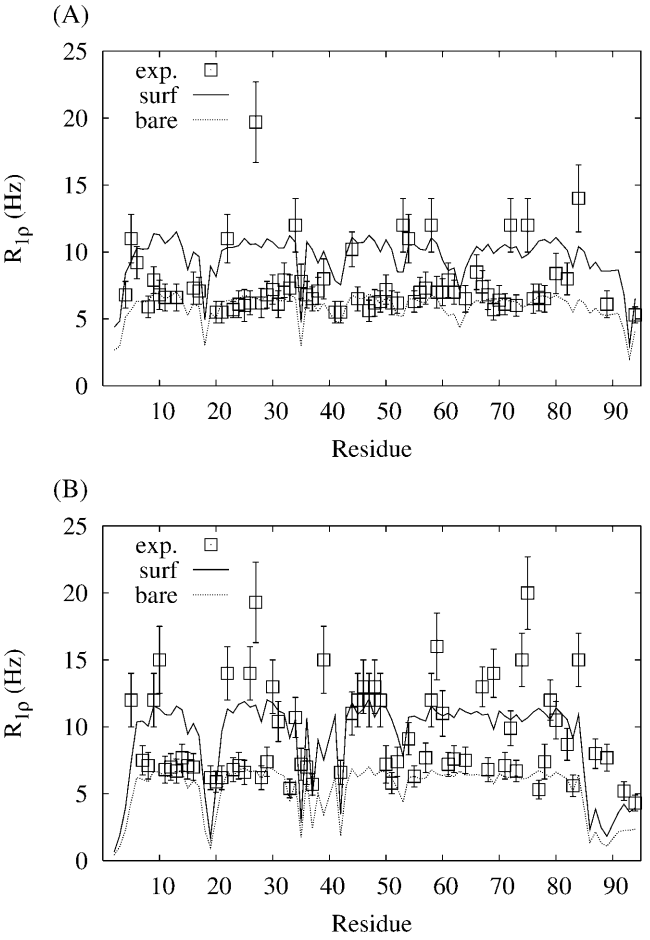


FIGURE 10 Same as Fig. 8 for the $R_{1\rho}$ NMR relaxation rates of ^{15}N at $\nu(^1\text{H}) = 600$ MHz and $\omega_e = 1870$ Hz for the reduced (*A*) and oxidized (*B*) forms.

are almost contained within the two sets of data computed with the two different diffusive models. Note, however, that for low frequencies ($R_{1\rho}$, Fig. 10) the curve for the bare model is lower than that for the surface model, whereas it is the opposite for high frequencies (R_1, Fig. 8). This behavior is expected, in that the bare model better describes the hydrodynamic properties of the high-rates tumbling produced by the protein backbone chain friction, whereas the surface model better describes the hydrodynamics of low-rates cooperative motions concerning the tumbling of the solvated molecule. The $R_{1\rho}$ relaxation rates are strongly influenced by the use of solvated or nonsolvated diffusion models. The choice of friction points located on the molecule or on the solvent shell can change $R_{1\rho}$ values by a factor 2, whereas the variation of R_1 relaxivities is only ~30%.

The extent of this effect can be physically related to the molecular solvation and surface. We have previously observed that these two latter parameters display by far the most evident differences between the statistics of the reduced and oxidized forms of Cyt b_5: the different charge density on the heme produces structural changes that modify the molecular surface: this produces a molecular expansion (Fig. 4) and a looser solvation shell in the oxidized form (Fig. 6).

The bare and surface diffusive models can be considered as lower and upper limits, respectively, for the $R_{1\rho}$ relaxation rates or, in general, for the low frequency contributions to spectral densities when the effects of the chemical shift modulation are ignored. Any frequency dependence of the NMR relaxation rates can then be related to a slow exchange between the leading terms in the different diffusion models. This kind of behavior is expected, for instance, when a molecular statistics allows the equilibrium between a basin of configurations A, with a first organized water shell, and another basin of configurations B, in which the first shell of water molecules is less defined. This is exactly the behavior suggested by Fig. 4, where, in the oxidized form, the molecule seems to return in a more compact form after the expansion.

Typical rotating frame experiments are performed in the fast exchange limit. Assuming k_a as the unique conversion rate from the basin of configurations A to the basin of configurations B, and k_b as the rate of the inverse process, the NMR spectrum displays a single peak for each ^{15}N nucleus when the amplitude of chemical shift modulation within the different conformations is smaller than $k = k_a + k_b$. Within this simplification, when ω_1 is larger than k and, therefore, the contribution from these conformational transitions to the $R_{1\rho}$ parameters of each ^{15}N nucleus is zero, the relaxation parameters tend to the population average for each nucleus. The MD simulations for the two forms of Cyt b_5 suggest that when the population of basin A is 100%, the asymptotic behavior of $R_{1\rho}$ would be the result of the bare model, whereas in the opposite case, when the population of basin A is zero, $R_{1\rho}$ would be the result of the surface model. These different ω_1 asymptotic behaviors of $R_{1\rho}$ can then be included in more complete models of $R_{1\rho}$ when chemical exchange events are significant, both in the fast and slow exchange limits (Trott and Palmer, 2002).

The MCD theory, in the present form, is not yet able to derive the protein dynamics related to infrequent events coupled with the molecular wobbling. Therefore, the RM2-II MCD calculations cannot obtain the timescale $1/k$ for the exchange between conformations revealed by the frequency dependence of $R_{1\rho}$ in the experiments. Nevertheless, the appearance of collective conformational exchange events in the oxidized form accounts for the larger number of ^{15}N nuclei whose $R_{1\rho}$ relaxation rates are frequency-dependent in the kHz region in the oxidized form with respect to the reduced form. Within this frame, the populations and the rate constant k within the two basins are the same for all the nuclei. Therefore, the contribution to $R_{1\rho}$ parameters due to chemical shift modulation depends only from the amplitude of this modulation, i.e., the difference in chemical shift of the nucleus in the two basins of conformations. This difference strongly depends on the average position and hydration of the nuclei within each basin.

CONCLUSIONS

Molecular dynamics simulations of 4.5 ns has been performed for the Fe(II) and Fe(III) redox forms of Cyt b_5, using as starting configurations the structures independently determined for each oxidation state. The two systems also differ for the point charges of iron and a few atoms of its ligands.

The configurational statistics obtained by MD simulations have been analyzed in detail. The most populated conformation of the histidine ligands has been found with the two imidazole planes parallel to each other and almost parallel to the B-D *meso* direction of the heme for both the oxidation states. The average dihedral angles describing this conformation are consistent with the solution NMR structures and with the pattern of chemical shifts due to the unpaired electron in the oxidized form.

One of the heme propionates is bent toward the iron ion in the oxidized form, whereas in the reduced form it is screened by a sodium ion. The bent conformation of the propionate in the oxidized form is strongly stabilized by hydrogen bond interactions with the backbone amidic H of Ser-64 and His-63. These interactions prevent the approach of sodium ions to the heme site and the screening of the heme propionate negative charge.

The distribution of secondary structure motives reveals that in both oxidation states the β-sheet structure is maintained, as well as the core of α-helices h3–h5. Two NMR-derived α-helices are found unstable. Helix 2 in both redox forms is rarely populated, thus showing its large propensity to unfolding as already suggested by NMR data in GdmCl 2 M solution. The mobility of this region is important because it contains His-39 bonded to Fe. Helix 6 is less

stable in the oxidized form because of the breaking of electrostatic interactions between the positively charged groups of the C-terminal and h6 region (residues 83–90) and negatively charged residues in h4 region (residues 53–60). In the oxidized form, the breaking of these interactions increases the stability of helix 4 and the solvent exposure of the region 53–60. This region has been identified as one of the negative patches in the recognition of Cyt *c* and modulation upon oxidation of the mobility and solvent accessibility has been already suggested by analyzing MD simulations of the bovine liver oxidized form.

The breaking of these electrostatic interactions and the related increase in molecular surface are indirect consequences of the larger negative charge close to Fe in the oxidized state: the larger positive charge in the oxidized state of iron is efficiently delocalized within the coordination site, whereas the negative charge of the bent propionate and the extrusion of a positive sodium ion from the active site has a stronger effect on the charge distribution in the protein. Negatively charged side chains at ~1.5 nm from Fe change conformation and positively charged side chains, which are in the reduced form involved in electrostatic interactions, are therefore affected.

The change in size and nature of the molecular surface occurring in the oxidized state has significant effects on the structure of the first solvation shell: in the reduced form the first water shell is relatively tightly bound to the protein, whereas in the oxidized state it moves with the molecular surface and it is, therefore, more mobile. The consequence is that, on average, there are ~50% more water molecules in the oxidized form than in the reduced one in the first solvation shell, but most of these molecules are less correlated to the solute.

The different structure of the solvation shell strongly affects the hydrodynamic properties of the molecule as suggested by reduction potential and NMR experiments. Mode-coupling diffusion theory has then been applied to calculate the molecular dynamics of H-N bonds in the protein backbone. The Smoluchowski diffusion equation, represented as the eigenvalue equation for the adjoint of the diffusion operator, has been solved by computing averages obtained along with the MD statistics. Two different diffusive models of the solute molecule have been used: a first bare model ignores the contribution of the solvation shell to the molecular friction, whereas a second surface model considers only the contribution of the first shell of water molecules correlated to the solute. The number of water molecules in this shell has been computed through the analysis of the solute-solvent radial distribution function and is not an adjustable parameter.

The results of the bare and surface diffusive models nicely match the lower and upper limits, respectively, of ^{15}N NMR $R_{1\rho}$ relaxation measurements on both the redox forms of the protein when chemical shift modulation contribution to NMR relaxation is not effective. This result suggests that a complete description of these relaxation parameters can be given in terms of an exchange process between conformational basins related to different molecular shapes. The MD statistics shows that the basin containing expanded molecules is more populated in the oxidized state because of the sampling of conformations where arrays of electrostatic interactions are broken. Solvation analysis shows that the sampling of these conformations produces a larger and looser first solvation shell in the oxidized form. Diffusion theory including the first shell contribution to the molecular relaxation modes allows a quantification of the effects of this difference onto the NMR relaxation parameters when chemical shift modulation is not effective. The range for R_1 and $R_{1\rho}$ NMR parameters obtained by diffusion theory matches the range obtained by measurements, without the need of adjustable parameters.

Several NMR experimental techniques allow the direct measurement of magnetic relaxation due to dipolar mechanisms, inhibiting the effects of the chemical shift modulation (Fushman et al., 1999). The interpretation of these data also requires accurate modeling of population-averaged rotational diffusion tensor components (Ghose et al., 2001) and, therefore, the improvement of diffusion theory here reported will be of great help.

The major limitation of the method is in the MD conformational sampling of compact and expanded molecular shapes: during ~5 ns of simulation the expansion of the oxidized form occurs only once and, therefore, no convergence of dynamics is expected within these statistics and the frequency dependence of NMR parameters is not captured. Nevertheless, the presence or absence of such expanded conformations is an indication of the occurrence of such dynamical process as it is clearly demonstrated by NMR parameters. Moreover, the distribution of the frequency dependence in the kHz region of $R_{1\rho}$ among the whole molecule can be better explained in terms of a global change of hydrodynamic properties in the molecule, rather than in terms of local barrier transitions.

The method here applied, i.e., the combination of computer simulations of the configurational statistics with diffusion theory to obtain the molecular dynamics, needs only configurational averages. Therefore, many computational methods based on umbrella sampling (Bartels and Karplus, 1998) or generalized-ensembles MD or Monte Carlo (Mitsutake et al., 2001), which allow more efficient sampling of molecular conformations, can be used and are expected to allow a better interpretation of NMR relaxation parameters related to slow conformational exchange processes that are of crucial importance in biological events.

SUPPLEMENTARY MATERIAL

An online supplement to this article can be found by visiting BJ Online at http://www.biophysj.org.

REFERENCES

Arnesano, F., L. Banci, I. Bertini, and I. C. Felli. 1998a. The solution structure of oxidized rat microsomal cytochrome b_5. *Biochemistry.* 37: 173–184.

Arnesano, F., L. Banci, I. Bertini, and D. Koulougliotis. 1998b. Solution structure of oxidized rat microsomal cytochrome b_5 in the presence of 2guanidinium chloride: monitoring the early steps in protein unfolding. *Biochemistry.* 37:17082–17092.

Arnesano, F., L. Banci, I. Bertini, D. Koulougliotis, and A. Monti. 2000. Monitoring mobility in the early steps of unfolding: the case of oxidized cytochrome b_5 in the presence of 2guanidinium chloride. *Biochemistry.* 39:7117–7130.

Banci, L., I. Bertini, C. Cavazza, I. C. Felli, and D. Koulougliotis. 1998a. Probing the backbone dynamics of oxidized and reduced rat microsomal cytochrome b_5 via ^{15}N rotating frame relaxation measurements: biological implications. *Biochemistry.* 37:12320–12330.

Banci, L., I. Bertini, P. Turano, and C. Luchinat. 1998b. The solution structure of redox proteins and beyond. Kluwer Academic Publishers, New York. 225–238.

Banci, L., I. Bertini, G. Cavallaro, and C. Luchinat. 2002. Chemical shift-based constraints for solution structure determination of paramagnetic low spin heme proteins with bis-His and His-Cn axial ligands. The cases of oxidized cytochrome b_5 and Met^{80}Ala cyano-cytochrome c. *J. Biol. Inorg. Chem.* 7:416–426.

Banci, L., I. Bertini, I. C. Felli, L. Krippahl, K. Kubicek, J. J. G. Moura, and A. Rosato. 2003. A further investigation of the cytochrome b_5-cytochrome *c* complex. *J. Biol. Inorg. Chem.* 8:777–786.

Banci, L., I. Bertini, F. Ferroni, and A. Rosato. 1997a. Solution structure of reduced microsomal cytochrome b_5. *Eur. J. Biochem.* 249:270–279.

Banci, L., G. G. Savellini, and P. Turano. 1997b. Molecular dynamics study in explicit water of the reduced and oxidized forms of yeast iso-1-cytochrome c. Solvation and dynamic properties of the two oxidation states. *Eur. J. Biochem.* 249:716–723.

Banci, L., and C. Presenti. 2000. Perspectives in inorganic structural biology. *J. Biol. Inorg. Chem. JBIC.* 5:422–431.

Bartels, C., and M. Karplus. 1998. Probability distributions for complex systems: adaptive umbrella sampling of the potential energy. *J. Phys. Chem. B.* 102:865–880.

Bayly, C. I., P. Cieplak, W. D. Cornell, and P. A. Kollman. 2002. A well-behaved electronics potential based method using charge restraints deriving atomic charges: the RESP model. *J. Phys. Chem.* 97:10269–10280.

Berendsen, H. J. C., J. P. M. Postma, W. F. van Gunsteren, A. Di Nola, and J. R. Haak. 1984. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* 81:3684–3690.

Bertini, I., H. B. Gray, S. J. Lippard, and J. S. Valentine, editors. 1994. Bioinorganic Chemistry. University Science Books, Mill Valley, CA.

Bertini, I., A. Sigel, and H. Sigel, editors. 2001. Handbook on Metalloproteins. Marcel Dekker, New York.

Brooks 3rd, C. L., M. Karplus, and B. M. Petitt. 1988. Proteins: a Perspective of Dynamics, Structure and Thermodynamics. J. Wiley and Sons, New York.

Canters, G. W., and M. Van de Kamp. 1992. Protein-mediated electron transfer. *Curr. Biol.* 2:859–869.

Case, D. 2002. Molecular dynamics and NMR spin relaxation in proteins. *Acc. Chem. Res.* 35:325–331.

Cavanagh, J., and W. J. Fairbrother, A. G. Palmer 3rd, and N. J. Skelton. 1996. Protein NMR Spectroscopy. Academic Press, San Diego, CA.

Chandrasekhar, I., G. M. Clore, A. Szabo, A. M. Gronenborn, and B. R. Brooks. 1992. A 500-ps molecular dynamics simulation study of interleukin-1β in water. Correlation with nuclear magnetic resonance spectroscopy and crystallography. *J. Mol. Biol.* 226:239–250.

Clore, G. M., A. Szabo, A. Bax, L. E. Kay, P. C. Driscoll, and A. M. Gronenborn. 1990. Deviations from the simple two-parameter model-free approach to the interpretation of nitrogen-15 nuclear magnetic relaxation of proteins. *J. Am. Chem. Soc.* 112:4989–4991.

Corin, A., G. McLendon, Q. Zhang, R. Hake, J. Falvo, K. S. Lu, R. B. Ciccarelli, and D. Holzschu. 1991. Effects of surface amino acid replacements in cytochrome-c peroxidase on complex formation with cytochrome c. *Biochemistry.* 30:11585–11595.

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

Dangi, B., J. Blankman, C. J. Miller, B. F. Volkman, and R. D. Guiles. 1998a. Contribution of backbone dynamics to entropy changes occurring on oxidation of cytochrome b_5; can redox-linked changes in hydrogen bond networks modulate reduction potentials? *J. Phys. Chem. B.* 102: 8201–8208.

Dangi, B., S. Sarma, C. Yan, D. L. Banville, and R. D. Guiles. 1998b. The origin of differences in the physical properties of the equilibrium forms of cytochrome b_5 revealed through high-resolution structures and backbone dynamic analyses. *Biochemistry.* 37:8289–8302.

de la Torre, J. G., M. Huertas, and B. Carrasco. 2000. HYDRONMR: prediction of NMR relaxation of globular proteins from atomic-level structures and hydrodynamic calculations. *J. Magn. Res.* 147:138–146.

Desvaux, H., N. Birlirakis, C. Wary, and P. Berthault. 1995. Study of slow molecular motions in solution using off-resonance irradiation in homonuclear NMR. II. Fast chemical exchange processes. *Mol. Phys.* 86:1059–1073.

Durley, R. C. E., and F. S. Mathews. 1996. Refinement and structural analysis of bovine cytochrome b_5 at 1.5 Ångstrom resolution. *Acta Crystallogr. D Biol. Crystallogr.* 52:65–72.

Eisenhaber, F., P. Lijnzaad, P. Argos, C. Sander, and M. Scharf. 1995. The double cubic lattice method: efficient approaches to numerical integration of surface area and volume and to dot surface contouring of molecular assemblies. *J. Comput. Chem.* 16:273–284.

Essmann, U., L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen. 1995. A smooth particle mesh Ewald method. *J. Chem. Phys.* 103:8577–8593.

Fausti, S., G. La Penna, C. Cuniberti, and A. Perico. 1999. Mode-coupling Smoluchowski dynamics of a double-stranded DNA oligomer. *Biopolymers.* 50:613–629.

Fausti, S., G. La Penna, C. Cuniberti, and A. Perico. 2000. Diffusive dynamics in a detailed potential: application to biological macromolecules. *Mol. Simul.* 24:307–324.

Fernandes, M. X., A. Ortega, M. C. López Martínez, and J. G. de la Torre. 2002. Calculation of hydrodynamic properties of small nucleic acids from their atomic structure. *Nucleic Acids Res.* 30:1782–1788.

Fushman, D., N. Tjandra, and D. Cowburn. 1999. An approach to direct determination of protein dynamics from 15N NMR relaxation at multiple fields, independent of variable ^{15}N chemical shift anisotropy and chemical exchange contributions. *J. Am. Chem. Soc.* 121:8577–8582.

Ghose, R., D. Fushman, and D. Cowburn. 2001. Determination of the rotational diffusion tensor of macromolecules in solution from NMR relaxation data with a combination of exact and approximate methods. Application to the determination of interdomain orientation in multidomain proteins. *J. Magn. Reson.* 149:204–217.

Hake, R., G. McLendon, A. Corin, and D. Holzschu. 1992. Redox-dependent molecular recognition in proteins: site-directed mutagenesis suggests that cytochrome-c oxidation state governs binding and recognition to cytochrome-c peroxidase. *J. Am. Chem. Soc.* 114:5442–5443.

HPCC Group. 2002. NWCHEM: a Computational Chemistry Package for Parallel Computers, Vers. 4.1. High Performance Computational Chemistry Group, Pacific Northwest National Laboratory, Richland, WA.

Ito, A., and R. Sato. 1968. Purification by means of detergents and properties of cytochrome b_5 from liver microsomes. *J. Biol. Chem.* 243:4922–4923.

Jorgensen, W. L., J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. J. Klein. 1983. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79:926–935.

Kabsch, W., and C. Sander. 1983. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers.* 22:2577–2637.

Kadish, K. M., K. M. Smith, R. Guilard, and R. Guiland, editors. 1999. The Porphyrin Handbook. Academic Press, Burlington, MA.

Kollman, P. A., and K. M. J. Merz. 1990. Computer modeling of the interactions of complex molecules. *Acc. Chem. Res.* 23:246–252.

Koradi, R., M. Billeter, and K. Wüthrich. 1996. MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graph.* 14:51–55.

La Penna, G., S. Fausti, A. Perico, and J. A. Ferretti. 2000a. Smoluchowski dynamics of the v_{nd}/NK-2 homeodomain from *Drosophila melanogaster*: second order maximum correlation approximation. *Biopolymers.* 54: 89–103.

La Penna, G., A. Perico, and D. Genest. 2000b. Dynamics of a double-stranded DNA oligomer: mode-coupling diffusion approach and reduced rigid fragment models. *J. Biomol. Struct. Dyn.* 17:673–685.

La Penna, G., D. Genest, and A. Perico. 2003. Modeling the dynamics of the solvated SL1 domain of HIV-1 genomic RNA. *Biopolymers.* 69:1–14.

La Penna, G., A. Mitsutake, M. Masuya, and Y. Okamoto. 2003. Molecular dynamics of C-peptide of ribonuclease A studied by replica-exchange Monte Carlo method and diffusion theory. *Chem. Phys. Lett.* 380:609–619.

La Penna, G., R. Pratolongo, and A. Perico. 1999. Mode-coupling Smoluchowski dynamics of polymers in the limit of rigid structures. *Macromolecules.* 32:506–513.

Lederer, F. 1994. The cytochrome b_5 fold: an adaptable module. *Biochimie.* 76:674–692.

Lipari, G., and A. Szabo. 1982. Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity. *J. Am. Chem. Soc.* 104:4546–4559.

Mathews, F. S. 1985. The structure, function and evolution of cytochromes. *Progr. Biophys. Mol. Biol.* 45:1–56.

McCammon, J. A., and S. C. Harvey. 1987. Dynamics of proteins and nucleic acids. Cambridge University Press, Cambridge, UK.

Mitsutake, A., Y. Sugita, and Y. Okamoto. 2001. Generalized-ensemble algorithms for molecular simulations of biopolymers. *Biopolym. Pept. Sci.* 60:96–123.

Moore, G. R., and G. W. Pettigrew. 1990. Cytochromes c; Evolutionary, Structural and Physicochemical Aspects. Springer-Verlag, Berlin, Germany.

Moore, G. R., G. W. Pettigrew, and N. K. Rogers. 1986. Factors influencing redox potentials of electron transfer proteins. *Proc. Natl. Acad. Sci. USA.* 83:4998–4999.

Palmer 3rd, A. G. 2001. NMR probes of molecular dynamics: overview and comparison with other techniques. *Annu. Rev. Biophys. Biomol. Struct.* 30:129–155.

Palmer 3rd, A. G., C. D. Kroenke, and J. P. Loria. 2001. NMR methods for quantifying microsecond-to-millisecond motions in biological macromolecules. *Meth. Enzymol.* 339:204–238.

Pastor, R. W., and M. Karplus. 1988. Parametrization of the friction constant for stochastic simulations of polymers. *J. Phys. Chem.* 92:2636–2641.

Peerey, L. M., H. M. I. Brothers, J. T. Hazzard, G. Tollin, and N. M. Kostic. 1991. Unimolecular and bimolecular oxidoreduction reactions involving diprotein complexes of cytochrome c and plastocyanin. Dependence of electron transfer reactivity on charge and orientation of the docked metalloproteins. *Biochemistry.* 30:9297–9304.

Peng, J. W., and G. Wagner. 1992. Mapping of spectral density function using heteronuclear relaxation measurements. *J. Magn. Reson.* 98:308–332.

Perico, A., and R. Pratolongo. 1997. Maximum correlation mode-coupling approach to the Smoluchowski dynamics of polymers. *Macromolecules.* 30:5958–5969.

Peter, C., X. Daura, and W. F. van Gunsteren. 2001. Calculation of NMR-relaxation parameters for flexible molecules from molecular dynamics simulations. *J. Biomol. NMR.* 20:297–310.

Prompers, J. J., and R. Brüschweiler. 2002. General framework for studying the dynamics of folded and nonfolded proteins by NMR relaxation spectroscopy and MD simulation. *J. Am. Chem. Soc.* 124: 4522–4534.

Qian, C., Y. Yao, K. Ye, J. Wang, W. Tang, Y. Wang, W. Wang, J. Lu, Y. Xie, and Z. Huang. 2001. Effects of charged amino-acid mutation on the solution structure of cytochrome b_5 and binding between cytochrome b_5 and cytochrome c. *Protein Sci.* 10:2451–2459.

Rodgers, K. K., T. C. Pochapsky, and S. G. Sligar. 1988. Probing the mechanisms of macromolecular recognition: the cytochrome b_5-cytochrome c complex. *Science.* 240:1657–1659.

Rose, M. E. 1957. Elementary Theory of Angular Momentum. John Wiley & Sons, New York.

Rotne, J., and S. Prager. 1969. Variational treatment of hydrodynamic interaction in polymers. *J. Chem. Phys.* 50:4831–4837.

Rovira, C., B. Schulze, M. Eichinger, J. D. Evanseck, and M. Parrinello. 2001. Influence of the heme pocket conformation on the structure and vibrations of the Fe-CO bond in myoglobin: a QM/MM density functional study. *Biophys. J.* 81:435–445.

Ryckaert, J.-P., G. Ciccotti, and H. J. C. Berendsen. 1977. Numerical integration of the Cartesian equations of motion with constraints: molecular dynamics of *n*-alkanes. *J. Comp. Phys.* 23:327–341.

Scott, R. A., and A. G. Mauk. 1996. Cytochrome c. Multidisciplinary Approach. University Science Books, Sausalito, CA.

Spatz, L., and P. Strittmatter. 1971. Form of cytochrome b_5 that contains an additional hydrophobic sequence of 40 amino acids. *Proc. Natl. Acad. Sci. USA.* 68:1042–1046.

Stocker, U., and W. F. van Gunsteren. 2000. Molecular dynamics simulation of hen egg white lysozyme: a test of the GROMOS96 force field against nuclear magnetic resonance data. *Proteins.* 40:145–153.

Storch, E. M., and V. Daggett. 1995. Molecular dynamics simulation of cytochrome b_5: implications for protein-protein recognition. *Biochemistry.* 34:9682–9693.

Trott, O., and A. G. Palmer, III. 2002. $R_{1\rho}$ Relaxation outside of the fast-exchange limit. *J. Magn. Reson.* 154:157–160.

van Gunsteren, W. F. 1993. Molecular dynamics studies of proteins. *Curr. Opin. Struct. Biol.* 3:277–281.

von Bodman, S. B., M. A. Schulder, D. R. Jollie, and S. G. Sligar. 1986. Synthesis, bacterial expression, and mutagenesis of the gene coding for mammalian cytochrome b_5. *Proc. Natl. Acad. Sci. USA.* 83:9443–9447.

Wagner, G. 1993. Relaxation and protein mobility. *Curr. Opin. Struct. Biol.* 3:748–754.

Wagner, G., S. Hyberts, and J. W. Peng. 1993. Study of Protein Dynamics. Macmillan Press, New York. 220–257.

Walker, M. C., and G. Tollin. 1991. Laser flash photolysis studies of the kinetics of electron-transfer reactions of *Saccharomyces* flavocytochrome b_2: evidence for conformational gating of intramolecular electron transfer induced by pyruvate binding. *Biochemistry.* 30:5546–5555.

Wu, F. J., J. H. Gan, Z. X. Xia, Y. H. Wang, W. H. Wang, L. L. Xue, Y. Xie, and Z. X. Huang. 2000. Crystal structure of recombinant trypsin-solubilized fragment of cytochrome b_5 and the structural comparison with Val^{61}His mutant. *Proteins Struct. Funct. Genet.* 40:249–259.

Wu, Y., Y. Wang, C. Qian, J. Lu, E. Li, W. Wang, Y. Xie, J. Wang, D. Zhu, Z. Huang, and W. Tang. 2001. Solution structure of cytochrome b_5 mutant (E44/48/56A/D60A) and its interaction with cytochrome c. *Eur. J. Biochem.* 268:1620–1630.

Xu, X. P., and D. A. Case. 2002. Probing multiple effects on ^{15}N, $^{13}C\alpha$, $^{13}C\beta$ and $^{13}C'$ chemical shifts in peptides using density functional theory. *Biopolymers.* 65:408–423.

Zhou, J. S., and N. M. Kostic. 1992. Photoinduced electron transfer from zinc cytochrome c to plastocyanin is gated by surface diffusion within the metalloprotein complex. *J. Am. Chem. Soc.* 114:3562–3563.

3.2 Paramagnetics-based restraints for Xplor-NIH
[*J. Biomol. NMR* 28, 249-261 (2004)]

Journal of Biomolecular NMR **28:** 249–261, 2004.
KLUWER/ESCOM

Paramagnetism-based restraints for Xplor-NIH

Lucia Banci[a], Ivano Bertini[a,*], Gabriele Cavallaro[a], Andrea Giachetti[a], Claudio Luchinat[b] & Giacomo Parigi[b]
[a]*CERM and Department of Chemistry and* [b]*CERM and Department of Agricultural Biotechnology, University of Florence via L. Sacconi, 6, I-50019 Sesto Fiorentino, Italy*

Received 24 June 2003; Accepted 6 October 2003

Key words: Curie-dipolar cross correlations, paramagnetism, pseudocontact shifts, relaxation rates, residual dipolar couplings, structure calculations, Xplor-NIH

Abstract

Modules that use paramagnetism-based NMR restraints have been developed and integrated in the well known program for solution structure determination Xplor-NIH; the complete set of such modules is called PARArestraints for Xplor-NIH. Paramagnetism-based restraints are paramagnetic relaxation enhancements, pseudocontact shifts, residual dipolar couplings due to metal and overall magnetic anisotropy, and cross correlation between Curie relaxation and nuclear-nuclear dipolar relaxation. The complete program has been tested by back-calculating NOEs and paramagnetism-based restraints from the X-ray structure of cytochrome c_{553} from *B. pasteurii*. Furthermore, the same experimental restraints previously used to determine the solution structure of cytochrome c_{553} itself, of cytochrome b_5, and of calbindin D_{9k} with the program PARAMAGNETIC DYANA, have been used for structure calculations by using PARArestraints for Xplor-NIH. The agreement between the two programs is quite satisfactory and validates both protocols.

Introduction

NMR spectroscopy is a well established technique for structural determination which flanks X-ray crystallography, and its use is steadily increasing over the years. Most of its applications are still devoted to non metal containing or diamagnetic metal ion containing proteins. This reflects an intrinsic difficulty in studying, through NMR, systems containing paramagnetic metal ions, which have profound effects in the NMR spectra, often determining severe line broadening and sizable reduction in the detectable constraints, particularly NOEs. However, when tailored experiments are developed and optimised for paramagnetic proteins, and signals affected by the paramagnetic center are detected, the paramagnetism-induced effects on NMR parameters are precious source of structural information. In particular, these new type of restraints are very useful to structurally define the region around the metal ion (Bertini et al., 2001a, 2002a, b). Several applications have been reported up to now (Tolman et al., 1995; Gochin and Roder, 1995; Huber et al., 1996; Banci et al., 1996, 1997, 1998a; Bertini et al., 1997; Bentrop et al., 1997; Bax and Tjandra, 1997; Turner et al., 1998; Dunham et al., 1998; Arnesano et al., 1998, 1999; Boisbouvier et al., 1999; Kechuan and Gochin, 1999; Hus et al., 2000; Barbieri et al., 2002). It was also shown that, in principle and in a few real cases, paramagnetism-based restraints provide enough information to obtain the fold of the protein backbone if used in conjunction with few other information, without any NOE restraints (Hus et al., 2000; Bertini et al., 2002b).

Paramagnetism-based restraints originate from the perturbations of the NMR parameters due to the coupling between the nuclear spin and the unpaired electron spin. The paramagnetic contributions to nuclear relaxation rates, the pseudocontact shifts, the residual dipolar couplings due to magnetic anisotropy of the paramagnetic molecule, and the cross correlations

*To whom correspondence should be addressed. E-mail: ivanobertini@cerm.unifi.it

between Curie and dipolar interactions depend on geometrical properties of the molecule which, once extracted, can be used in structural calculations (Bertini et al., 2001b, 2002a, b). Relaxation rate, pseudocontact shift and cross correlation restraints contain information on the distance of the metal ion from the resonating nuclei. Pseudocontact shift values also depend on the orientation of the metal nucleus vector in the magnetic susceptibility frame. Cross correlation values provide information, e.g., on the angle between the metal-nucleus direction and a nucleus-nucleus dipole direction. Self-orientation residual dipolar coupling values provide information on the orientation of dipole of the two coupled nuclei in the molecular magnetic susceptibility frame.

The amount of information provided by these restraints can be so large that diamagnetic proteins containing a metal binding site may be conveniently investigated by substituting the diamagnetic metal ion with a paramagnetic one (Bertini et al., 2001c). Furthermore, it may be convenient to substitute different paramagnetic metal ions in the same binding site, in order to have several sets of data, which are often complementary (Bertini et al., 2001a, d). Indeed, the metal susceptibility tensor depends on the nature and the coordination properties of the metal ion and therefore different metal ions provide independent information.

The X-PLOR package (Clore et al., 1985), derived from the program CHARMM, and its following implementations (CNS) (Brunger et al., 1998) is one of the most popular programs for obtaining protein solution structures through structural restraints, simulated annealing calculations and energy minimization. The Xplor-NIH program (Schwieters et al., 2003) is a version which contains all the functionality present in the last release of X-PLOR, and incorporates new features as the modules for torsion angle dynamics, a C++ framework and the interfaces with Python and TCL, and additional restraints for structure refinement. Biomolecular solution structure determination is achieved by minimizing a target function calculated by adding a term related to experimental NMR restraints to the terms related to covalent geometry and non-bonded interactions. Minimization procedures comprise molecular dynamics in Cartesian and torsion angle spaces, and conventional gradient-based minimization.

Residual dipolar coupling restraints due to molecular magnetic anisotropy and/or to induced molecular orientation were included in the program Xplor-NIH, and their efficiency tested (Tjandra et al., 1997, 2000; Clore et al., 1998; Clore and Garrett, 1999). Their use was largely demonstrated to be relevant to solve structural calculation problems (Tjandra et al., 1997; Clore et al., 1999; Clore, 2000; Chou et al., 2000; Sass et al., 2001; Clore and Bewley, 2002; de Alba and Tjandra, 2002; Clore and Schwieters, 2003). Pseudocontact shift restraints were also included in the program Xplor-NIH and used to refine the structures of cytochrome *c* and its mutant L94V (Gochin and Roder, 1995), to position the monomeric subunits within a dimer (Gaponenko et al., 2002), and to obtain the structure of a DNA octamer complexed to chromomycin-A_3 (Tu and Gochin, 1999; Gochin, 2000). The pseudocontact shifts module is, however, not distributed as a documented routine. Finally, also paramagnetic enhancements to relaxation rates were included as such as restraints in the CNS package (Donaldson et al., 2003). However, a single program package based on Xplor-NIH which permits the integrated use of all the paramagnetism-based restraints does not exist. In our experience with the programs Diana (Güntert et al., 1991), Dyana (Güntert et al., 1997) and Cyana (Herrmann et al., 2002), only such integration permits an efficient use of such restraints by non specialists.

We have now included all the paramagnetism-based restraints into the program Xplor-NIH in a uniform way and by properly considering all their interconnections. The whole set of modules which allows the use of paramagnetic restraints is called PARArestraints for Xplor-NIH. We have tested the efficiency of the protocol on an already determined solution structure (cytochrome c_{553} from *B. pasteurii*) using simulated values of pseudocontact shifts, self-orientation residual dipolar couplings and Curie-dipolar cross correlations. Then, three protein structures have been recalculated with PARArestraints for Xplor-NIH by using the same set of experimental restraints used with the analogous PARAMAGNETIC DYANA program (Güntert and Wüthrich, 1991; Güntert et al., 1997; Banci et al., 1998b; Bertini et al., 2002a). The results of the two approaches, i.e. PARAMAGNETIC DYANA and PARArestraints for Xplor-NIH, are also compared. Although the tests are made on ^{1}H data, the program is suitable for heteronuclei as well.

Program implementation

The paramagnetic package implemented in Xplor-NIH consists in the algorithms XDIPO_PCS, XDIPO_RDC, XANGLE, XCCR and XT1DIST. The routines for the introduction of pseudocontact shift and residual dipolar coupling restraints are modifications of the existing XDIPO routine (Tjandra et al., 2000).

Restraints have been implemented in the structure calculations by using the typical least square energy penalty:

$$E = \sum_l w_l \sum_i [\max(|X_{\mathrm{i,obs}} - X_{\mathrm{i,calc}}| - \mathrm{tol}_i, 0)]^2, \quad (1)$$

where the index l runs over all classes of restraints, the index i on all experimental data of each class; tol_i indicates the tolerance on the ith restraint, and w_l the force constant of each class of restraints. Specific w_l values need to be defined whenever restraints of different nature are used together in structural calculations. The choice of the force constants is critical for a fruitful use of all restraints, since it dramatically influences the convergence of the calculations. The optimal force constant for each class of restraints must be found in order to make that restraint effective in structure calculations without an unreasonable increase of the energy for the other restraints. Some results and guidelines on this will be presented later.

The contributions to the energy gradient from each class of restraints, needed to integrate the equations of motion, are calculated as the first derivative of the energy terms, E, with respect to the Cartesian coordinates.

Inclusion of pseudocontact shift restraints

The presence of a paramagnetic metal ion induces a shift on the nuclear resonances. This shift is determined by two contributions: A contact contribution, due to through-bond nuclear spin electron spin coupling, and a pseudocontact shift contribution. The pseudocontact term is due to the dipolar interaction between a nuclear magnetic moment and an average induced electron magnetic moment. The latter depends on the scalar product of the metal magnetic susceptibility tensor with the applied magnetic field vector. As a result, the pseudocontact shift values depend on the position of each observed nucleus in the magnetic metal susceptibility frame, with origin on the metal ion, and on the anisotropy of the latter, according to the following equation (Kurland and McGarvey, 1970; Bertini et al., 2001b, 2002a):

$$\begin{aligned}\delta_i^{\mathrm{pcs}} = {} & \frac{1}{12\pi r_i^3}[\Delta\chi_{\mathrm{ax}}(3\ \cos^2\ \vartheta_i - 1) + \\ & \tfrac{3}{2}\Delta\chi_{\mathrm{rh}} \sin^2\ \vartheta_i\ \cos\ 2\varphi_i],\end{aligned} \quad (2)$$

where r_i is the distance between the atom i and the metal ion, ϑ_i and φ_i are the polar angles of atom i with respect to the principal axes of the metal magnetic susceptibility tensor centered on the metal ion, and

$$\begin{aligned}\Delta\chi_{ax} &= \chi_{zz} - \tfrac{\chi_{xx}+\chi_{yy}}{2}, \\ \Delta\chi_{rh} &= \chi_{xx} - \chi_{yy}.\end{aligned} \quad (3)$$

In order to introduce such restraints in the calculation of the structure, a pseudoresidue has to be defined, which describes the orientation and the origin of the metal susceptibility tensor. The latter in general coincides with the position of the metal ion (Banci et al., 1996). Furthermore, the magnetic anisotropy values, $\Delta\chi_{\mathrm{ax}}$ and $\Delta\chi_{\mathrm{rh}}$, must be obtained. They can be obtained with the module FRUN in an iterative fashion. FRUN calculates, through a best fit procedure, the values of the anisotropic part of the metal susceptibility tensor from the measured pseudocontact shifts and the available protein structure as inputs. In the first cycle the tensor parameters can be estimated either theoretically or from a preliminary protein structure obtained using other restraints. In the latter case, a fit is done over the five parameters $\chi_{zz} - \overline{\chi}$, $\chi_{xx} - \chi_{yy}$, χ_{xy}, χ_{xz} and χ_{yz}, as δ^{pcs} depends linearly on such parameters in any arbitrary reference frame (Kemple et al., 1988), and it does not depend on the trace of the magnetic susceptibility tensor. A diagonalization of the anisotropic part of the magnetic susceptibility tensor is then performed to obtain the principal values of the tensor and to calculate the anisotropy values in Equation 3.

The algorithm XDIPO_PCS, adapted from the existing algorithm XDIPO, applies pseudocontact shift restraints in structural calculations, using, in addition to the latter, also the values of $\Delta\chi_{\mathrm{ax}}$ and $\Delta\chi_{\mathrm{rh}}$ as input parameters. No assumption on the position of the metal, and thus on the origin of the tensor, is needed.

In practical applications the following protocol is suggested: (i) Calculate N preliminary structures either without the inclusion of pseudocontact shift restraints, or by including pseudocontact shift restraints and using theoretical estimates for the metal susceptibility anisotropies, (ii) on each structure of a subset characterized by the lowest global energy, calculate the values of the metal susceptibility anisotropies by fitting the experimental pseudocontact shift values with FRUN, then average the anisotropies, (iii) calculate N new structures including pseudocontact shifts

and the new average metal susceptibility anisotropy values, and so on until convergence is reached. The values of the magnetic susceptibility anisotropies is kept constant during the structure calculations (Banci et al., 1996).The scheme is summarized in Figure 1.

Errors in the values of the magnetic susceptibility anisotropies can be estimated through the bootstrap Monte Carlo method (Press et al., 1988), which consists in calculating the standard deviation of the different values obtained for the anisotropies after multiple removal of about 35% of randomly selected pseudocontact shifts.

Inclusion of residual dipolar coupling restraints

Self-orientation residual dipolar couplings (rdc) are restraints of the same kind of the residual dipolar coupling produced by the presence of an external orienting agent. In paramagnetic molecules, protein partial self-orientation in a magnetic field is induced by the magnetic anisotropy of the electron magnetic moment, as well as of the diamagnetic frame. As in the case of externally induced partial orientation, its effect on dipolar couplings depends on the coupled nuclei vector orientation within the magnetic susceptibility tensor and on the size of its anisotropy. The algorithm to include these restraints in structural calculations is XDIPO_RDC, also adapted from XDIPO. Residual dipolar coupling values are provided by the following equation, written for the X-H coupled nuclei (Bertini et al., 2001b, 2002a; Banci et al., 1998a):

$$\Delta\nu_{\mathrm{RDC}}(\mathrm{Hz}) = -\frac{1}{4\pi}\frac{B_0^2}{15kT}\frac{\gamma_{\mathrm{X}}\gamma_{\mathrm{H}}\hbar}{2\pi r_{\mathrm{XH}}^3}[\Delta\chi_{\mathrm{ax}}^{\mathrm{mol}}(3\cos^2\theta - 1) + \tfrac{3}{2}\Delta\chi_{\mathrm{rh}}^{\mathrm{mol}}\sin^2\theta\cos 2\Omega], \quad (4)$$

where θ is the angle between the X-H vector and the z axis of the χ^{mol} tensor, Ω is the angle which describes the position of the projection of the X-H vector on the xy plane of the χ^{mol} tensor, relative to the x axis, and $\Delta\chi_{\mathrm{ax}}^{\mathrm{mol}}$ and $\Delta\chi_{\mathrm{rh}}^{\mathrm{mol}}$ are defined as

$$\Delta\chi_{\mathrm{ax}}^{\mathrm{mol}} = \chi_{zz}^{\mathrm{mol}} - \frac{\chi_{xx}^{\mathrm{mol}} + \chi_{yy}^{\mathrm{mol}}}{2}, \qquad \Delta\chi_{\mathrm{rh}}^{\mathrm{mol}} = \chi_{xx}^{\mathrm{mol}} - \chi_{yy}^{\mathrm{mol}}, \quad (5)$$

analogously to Equation 3, where the magnetic molecular susceptibility anisotropy tensor is the sum of the diamagnetic and the metal magnetic susceptibility tensors.

The module FRUN can be again used for obtaining $\Delta\chi_{\mathrm{ax}}^{\mathrm{mol}}$ and $\Delta\chi_{\mathrm{rh}}^{\mathrm{mol}}$ from fitting the experimental rdc to the available structure. Structure calculations and updates of the magnetic susceptibility anisotropies are performed iteratively, in a similar fashion to that described for the pseudocontact shift restraints.

Experimental residual dipolar couplings can be obtained either by performing measurements at two different fields or by performing measurements on the paramagnetic and diamagnetic samples at a single field. In the latter case, residual dipolar couplings, as obtained by subtracting the ^1J of the diamagnetic species from the ^1J of the paramagnetic species, only depends on the paramagnetic metal ion contribution to the magnetic susceptibility tensor that is the same which determines pseudocontact shifts. Therefore, in this case the magnetic susceptibility anisotropy tensor obtained from pseudocontact shifts can be used in Equation 4. Local motions can alter the measured residual dipolar coupling values with respect to what calculated from Equation 4, the resulting effect being that of obtaining smaller values of $\Delta\chi_{\mathrm{ax}}$ and $\Delta\chi_{\mathrm{rh}}$ (Tolman et al., 1997; Bertini et al., 2001c). The use of $\Delta\chi_{\mathrm{ax}}$ and $\Delta\chi_{\mathrm{rh}}$ values obtained from the pseudocontact shift restraints actually evidenced the effects of internal mobility on residual dipolar couplings (Barbieri et al., 2002).

A possible contribution to the difference between the ^1J values of the paramagnetic and diamagnetic species due to the dynamic frequency shift ($\Delta\nu_{\mathrm{DFS}}$) should be also taken into account. The $\Delta\nu_{\mathrm{DFS}}$ contribution to ^1J, due to cross correlation between the dipole-dipole relaxation and the Curie relaxation originating from the coupling of the static magnetic moment of the unpaired electron and the nuclear spin is given by (Bertini et al., 2002b):

$$\Delta\nu_{\mathrm{DFS}} = \frac{\mu_0}{4\pi}\frac{3B_0\gamma_{\mathrm{H}}\gamma_{\mathrm{X}}\hbar\chi}{20\pi^2}\left(\frac{\gamma_{\mathrm{H}}}{r_{\mathrm{HX}}^3 r_{\mathrm{HS}}^3}\frac{3\cos^2\theta_{\mathrm{SHX}} - 1}{2}\frac{\omega_I\tau_{\mathrm{r}}^2}{1+\omega_I^2\tau_{\mathrm{r}}^2} + \frac{\gamma_{\mathrm{X}}}{r_{\mathrm{HX}}^3 r_{\mathrm{XS}}^3}\frac{3\cos^2\theta_{\mathrm{SXH}} - 1}{2}\frac{\omega_X\tau_{\mathrm{r}}^2}{1+\omega_X^2\tau_{\mathrm{r}}^2}\right), \quad (6)$$

where the angle θ_{Sij} (i, j = H, X) is that between the ij axis and the i-metal ion axis, r_{iS} is the i-metal ion distance, the correlation time, τ_{r}, is determined by the reorientation of the two vectors and

$$\chi = \mu_0\mu_B^2 g_J^2\frac{S(S+1)}{3kT},$$

or

$$\chi = \mu_0\mu_B^2 g_J^2\frac{J(J+1)}{3kT},$$

```
flags exclude * include bonds angle impr vdw noe xpcs end

PARAMETERS, STRUCTURE, COORDINATES, TOPOLOGY, NOE, DIHEDRALS

xdipo_pcs
 nres=500
 potential square
 coeff 594.925 -32.362          # tensor parameters
 @PCS.tbl                       # Input file
end

      Cycle on the iterations for the tensor parameters

   Cycle on the number of structures for each iteration
  coor swap end
  coor copy end
  noe
   potential * soft
  end

              Initial minimization

            High-temperature dynamics

                  Cooling

  noe
   potential * square
  end

              Final minimization

  xdipo_pcs
   son                                  # Save tensor parameters for averaging
   evaluate ($filename="1pse"+encode($count1)+"_"+encode($count))
   save $filename                       # Output file containing experimental and calc pcs
   frun 1                               # Perform FRUN on metal 1
  end

          Write out the minimized structure

 xdipo_pcs
  fmed 20 1                             # Perform the averaging of the tensor parameters
                                       # for metal 1 on the best 20% of the structures
  coeff $chiax $chirh                  # Update tensor parameters
  fmed 1 0                              # Reset stored tensor parameters
 end

xdipo_pcs
 erron 2000 35              # Monte Carlo error estimate: 2000 FRUN calculations
                            # randomly discarding 35% of the data
 frun 1
end
```

Figure 1. Scheme of the protocol for including pcs restraints in the structure calculation.

for lanthanides and actinides (g_J is the g electron factor in lanthanides and actinides). Such contribution to ^1J is small (with respect to residual dipolar coupling values) and decreases with the third power of the distance between the observed nuclei and the metal ion. The second term in Equation 6, shown to be present for the diamagnetic case (Werbelow, 1996), was derived for the paramagnetic case by H. Desvaux (pers. commun.), who also predicted a third smaller contribution (H. Desvaux, pers. commun.). In any case, the overall paramagnetic dynamic frequency shift to ^1J is expected to be negligible, and can be safely not taken into account in the structural calculations.

The module XANGLE was also implemented to use as restraints in structure calculations the polar θ and ϕ angles describing the orientation of the vector connecting a pair of coupled nuclear spins with respect to an arbitrary reference frame. This information can be straightforwardly introduced in structure calculation algorithms, thus making the use of the residual dipolar couplings restraints more efficient, as otherwise they are difficult to handle due to the complicated form of the corresponding energy surface,

which causes large degeneracy in the solutions. Such restraints may result useful when several sets of self-orientation residual dipolar couplings are available, as obtained from measurements on the same molecule when different paramagnetic metal ions are alternatively bound to the same binding site. Equation 4 can be written in the general form, valid in any reference system (Moltke and Grzesiek, 1999; Barbieri et al., 2002) as

$$\Delta\nu_{\mathrm{RDC}}(\mathrm{Hz}) = -\frac{1}{4\pi}\frac{B_0^2}{15kT}\frac{\gamma_{\mathrm{X}}\gamma\mathrm{H}\hbar}{2\pi r_{\mathrm{XH}}^3}\left[\frac{\chi_{zz}-\overline{\chi}}{2}(3\cos^2\theta-1) + \frac{\chi_{xx}-\chi_{yy}}{2}\sin^2\theta\cos 2\phi + \chi_{xy}\sin^2\theta\sin 2\phi + \chi_{xz}\sin 2\theta\cos\phi + \chi_{yz}\sin 2\theta\sin\phi\right]. \quad (7)$$

If the magnetic susceptibility anisotropy tensors can be calculated from the pseudocontact shifts, the values of residual dipolar couplings obtained on systems containing different metal ions (>2) can provide the orientations of the internuclear vectors, in terms of θ and ϕ angles. Having these experimental data, the following energy penalty term can be added to the global energy penalty in the structural calculations (Barbieri et al., 2002)

$$E_{\mathrm{angles}} = w_{\mathrm{angles}}\sum_i[1-(\mathbf{u}_{\mathrm{i}}\cdot\mathbf{v}_{\mathrm{i}})^2], \quad (8)$$

where w_{angles} is the force constant for this class of restraints, the $\mathbf{u}_{\mathrm{i}}$ vector has coordinates ($\sin\theta_i\cos\phi_i$, $\sin\theta_i\sin\phi_i$, $\cos\theta_i$) and

$$\mathbf{v}_{\mathrm{i}} = \frac{(\mathbf{r}_{\mathrm{H}}-\mathbf{r}_{\mathrm{X}})_i}{|\mathbf{r}_{\mathrm{H}}-\mathbf{r}_{\mathrm{X}}|_i}, \quad (9)$$

where $\mathbf{r}_{\mathrm{X}}$ and $\mathbf{r}_{\mathrm{H}}$ are the coordinate vectors of the X and H atoms, defined in any external reference system. This restraint permits two equivalent minima, corresponding to the two possible orientations (0° and 180°) of $\mathbf{v}_{\mathrm{i}}$ with respect to $\mathbf{u}_{\mathrm{i}}$.

Inclusion of restraints derived from cross correlations between Curie and dipolar relaxation

In a paramagnetic molecule the two components of a spin doublet may experience a difference in linewidth due to cross correlation between the nuclear dipole-dipole relaxation and Curie relaxation, originating from dipolar coupling between the nuclear spin and the static time-averaged electron magnetic moment. For the two components of the proton spin doublet in a dipole-dipole coupled HX system, the difference in linewidth, calculated in the assumption of isotropic χ tensor, is given by (Bertini et al., 2002a)

$$\Delta(\Delta\nu_{1/2}) = \frac{\mu_0}{4\pi}\frac{B_0\gamma_{\mathrm{H}}^2\gamma_{\mathrm{X}}\hbar\chi}{10\pi^2 r_{\mathrm{HS}}^3 r_{\mathrm{HX}}^3}\frac{3\cos^2\theta_{\mathrm{SHX}}-1}{2}\left(4\tau_{\mathrm{r}}+\frac{3\tau_{\mathrm{r}}}{1+\omega_I^2\tau_{\mathrm{r}}^2}\right) = \frac{3\cos^2\theta_{\mathrm{SHX}}-1}{r_{\mathrm{HS}}^3}k_{\mathrm{CCR}}, \quad (10)$$

where the symbols have the same meaning as in Equation 4. All terms not depending on the protein structure can be collected in the constant k_{CCR}. This contribution takes this form when the electron spin relaxation is fast with respect to the rotational time, τ_r.

This contribution to transverse relaxation contains structural information in terms of distances and angles between two vectors. These restraints can be included in structure calculations through a specific module (XCCR). They can be applied with a constant weighting factor, or the latter can be proportional to r_{HS}^3 times a constant weighting factor, in such a way that also nuclei far from the metal, and therefore characterized by small cross-correlation values, can have a contribution to the penalty energy. This latter approach is recommended.

This module requires the value of the constant k_{CCR} as input. The module FANTACCR has been developed, analogously to those for the restraints previously described, for estimating the constant k_{CCR} from experimental data and available structures, through best fit calculations to the experimental $\Delta(\Delta\nu_{1/2})$ data.

Inclusion of relaxation rate restraints

The experimental relaxation rates of nuclear spins coupled with unpaired electron spins are the sum of a diamagnetic and a paramagnetic contributions. The paramagnetic contribution is dominated (with the possible exception of nuclei separated by a few chemical bonds from the metal ion) by the dipolar coupling between the nuclear spin and the electron spin. The dipolar contribution is proportional to the inverse of the sixth power of the nuclear spin–metal ion (unpaired electron) distance (see Equation 11), and thus it is small for nuclei at large distance from the metal ion. Diamagnetic contributions can be evaluated by performing measurements on the diamagnetic analog of the molecule, or upper limit values can be estimated by

taking the average of the experimental relaxation rates of the paramagnetic molecule that are below a given threshold value. Furthermore, since for nuclei close to the paramagnetic center the diamagnetic contribution is small with respect to the paramagnetic contribution, the assumption of an upper limit value for the diamagnetic contribution produces a very small error on the nuclear spin-unpaired electron distances.

The paramagnetic contribution to nuclear relaxation rates can thus be used to obtain distance restraints between the observed nuclei and the metal ion. Actually, distances are usually used in structural calculations as upper distance limits, as a consequence of the overestimation of the diamagnetic contribution deriving from the use of the second approach described above and the consequent underestimation of the paramagnetic enhancement. A module (XT1DIST) was written to convert the rates into distances. With good approximation, for nuclei not directly coordinated to the metal ion, the relation between the paramagnetic contribution to the relaxation rate R_{1M} and the metal-nucleus distance r is (Bertini et al., 2001b)

$$R_{1M} = k/r^6, \tag{11}$$

where k is a constant. Distances can be generated from relaxation rates in two different ways. If the correlation time τ_c that modulates the nuclear spin–unpaired electron coupling is known, the constant k can be calculated from the Solomon equation

$$k = \frac{2}{15}\left(\frac{\mu_0}{4\pi}\right)^2 \gamma_I^2 \mu_{\text{eff}}^2 \left[\frac{7\tau_c}{1+\omega_s^2\tau_c^2} + \frac{3\tau_c}{1+\omega_I^2\tau_c^2}\right], \tag{12}$$

where $\mu_{\text{eff}}^2 = g_e^2\mu_B^2 S(S+1)$ or $\mu_{\text{eff}}^2 = g_J^2\mu_B^2 J(J+1)$ for lanthanides and actinides, μ_B is the electron Bohr magneton, γ_I is the proton magnetogyric ratio, g_e is the so-called free electron g value, ω_S is the electron Larmor frequency, ω_I is the proton Larmor frequency and S is the electron spin quantum number. τ_c is given by the sum of the rotational correlation rate, the exchange rate and the electron relaxation rate

$$\tau_c^{-1} = \tau_r^{-1} + \tau_M^{-1} + \tau_s^{-1} \tag{13}$$

and therefore it is essentially dominated by the fastest process. The estimated nuclear spin–metal ion distances can then be used as upper distance limits in structural calculations including a tolerance of 1 Å, which is added to the value of r. If a protein structure with good accuracy is already available, calculated for instance using other restraints, an upper limit value for the constant k can be calculated from the relaxation rates as a function of the distance r. Once a k value is obtained, Equation 11 is used again for obtaining r from the values of R_{1M}. In this way it is possible to adjust the distance restraints related to measurements of relaxation rates in an iterative fashion.

Structure calculations

An *ab initio* simulated annealing protocol was first applied performing 12 000 steps at high temperature (1000 K) and 6000 steps during cooling to 100 K with temperature intervals of 50 K. At each temperature, 333 steps of molecular dynamics simulation were performed with a time step of 5 fs. The resulting structures were then refined with a Powell minimization. Energy minimizations were then performed for 2000 steps each. Upper distance restraints from NOE and relaxation rate measurements were applied with a force constant of 209 kJ mol^{-1} Å^{-2} (50 kcal mol^{-1} Å^{-2}) during the whole calculation. Pseudocontact shifts, residual dipolar couplings and cross correlations between Curie and dipolar interactions were applied with force constants adjusted to have comparable contributions to the global energy.

In order to perform structure calculations of heme proteins, both *b*-type and *c*-type hemes were added to the Xplor-NIH library. Special patch residues, which are required to establish covalent linkages between the *c*-type heme and the cysteine residues which are bound to its vinyl substituents, were added as well. Ligands to the metal ions are provided as structural information, by the addition of upper distance limits between the ligand nuclei and the metal ion.

Results and discussion

Overall strategy in the use of paramagnetic restraints

The paramagnetism-based restraints are of different nature and have different geometric properties than diamagnetic restraints. For this reason they significantly contribute to increase the accuracy of the structure in addition to its precision. This property is based on the fact that these restraints have different dependences on the distance (r^{-3} for pcs and ccr, r^{-6} for relaxation rates) and on angular properties (tensor orientations for pcs and rdc, angles between vectors for ccr).

In order to be particularly effective, these restraints need to be applied since the early steps of the struc-

Table 1. Energetic and structural parameters for the family of structures of cytochrome c_{553} calculated with simulated data. Calculations are performed with Xplor-NIH using distance restraints only or distance restraints and paramagnetism-based restraints

	NOE only	NOE + paramagnetic restraints
Total energy (10^3 J mol^{-1})	280.3 ± 0.0	283.3 ± 0.4
(Energy NOE, pcs+rdc+ccr)	(0.00,−)	(0.04, 2.64)
BB RMSD to the mean (Å)	0.39 ± 0.04	0.29 ± 0.03
HA RMSD to the mean (Å)	0.62 ± 0.04	0.52 ± 0.04
$\Delta\chi_{ax}$ from pcs (10^{-32} m^3)	2.10 ± 0.09	2.21 ± 0.03
$\Delta\chi_{rh}$ from pcs (10^{-32} m^3)	−0.21 ± 0.07	−0.18 ± 0.04
$\Delta\chi_{ax}$ from rdc (10^{-32} m^3)	1.73 ± 0.15	2.18 ± 0.09
$\Delta\chi_{rh}$ from rdc (10^{-32} m^3)	−0.18 ± 0.10	−0.16 ± 0.05
BB RMSD minimized X-ray/Xplor-NIH (Å)	0.47 ± 0.05	0.35 ± 0.04
HA RMSD Minimized X-ray/Xplor-NIH (Å)	0.81 ± 0.07	0.72 ± 0.04

Table 2. Order of magnitude for typical absolute values of the axial magnetic metal susceptibility tensor for some common metal ions

Fe(III) (HS and LS)	3×10^{-32}
Fe(II) HS	2×10^{-32}
Co(II) HS	5×10^{-32}
Ce(III), Nd(II), Eu(II)	2×10^{-32}
Pr(III)	3×10^{-32}
Sm(III)	2×10^{-33}
Tb(III), Dy(III)	3×10^{-31}
Ho(III),Tm(III)	2×10^{-31}
Er(III),Yb(III)	1×10^{-31}

tural calculations and the force constants used for their inclusion in the energy penalty must be properly selected. For this purpose we have performed a series of test structural calculations to calibrate the weight of each class of restraints in such a way that they have a comparable contribution to the penalty energy with respect to the 'standard' restraints since the beginning of the calculations. It results that force constants of 21 kJ mol^{-1} ppm^{-2} for pseudocontact shifts, 21 kJ mol^{-1} Hz^{-2} for residual dipolar couplings and $r_{HS}^3 \times 4.2 \times 10^{-3}$Å^{-3} J mol^{-1} Hz^{-2} (with r in Å) for cross correlations are appropriate in most cases for values of the axial magnetic susceptibility anisotropy of the order of 2×10^{-32} m^3. Force constants for pcs and rdc should be set related to $\Delta\chi_{ax}$, their values being proportionally lower for larger $\Delta\chi_{ax}$.

In order to use pcs and rdc from the beginning in the structure calculation procedure, an estimation of the magnetic susceptibility tensor anisotropies is needed. Table 2 reports the typical values of $\Delta\chi_{ax}$ for some metal ions. We will show later that such estimation is generally enough for ensuring the convergence of the protocol to the correct value.

All the paramagnetism-based restraints are related to the metal ion, which constitutes the origin of each class of interactions. Therefore, the coordinates of the latter can be left free to vary and to be optimised during the structural calculations. In such a way the position of the metal ion can be carefully determined on the basis of experimental data without any assumption. Paramagnetic restraints therefore represent the unique way to locate an NMR silent metal ion within the molecular frame.

A comment is needed on the tolerance which should be used for each class of restraints in Equation 1. This strongly depends on the error in determining the experimental data, which mainly resides in the comparison with the diamagnetic values. In the case of pseudocontact shifts, if the experimental shift values for a corresponding diamagnetic molecule are available, then pcs with relatively high accuracy can be determined and low tolerance can be used. On the contrary, if only an estimate of the diamagnetic values can be obtained, larger tolerance should be used. It had been already verified and tested that it is appropriate to use a tolerance proportional to the value itself (10%) down to a lower limit which can be reasonably set between 0.1 and 0.3 ppm (Bertini et al., 2002b). Typical fixed tolerance values for residual dipolar couplings are 0.1–0.3 Hz, and for cross correlations are 0.1–0.2 Hz.

Structure calculations using simulated data

The package was tested with structural restraints calculated from the X-ray structure of the protein *B. pasteurii* cytochrome c_{553} (1C75) (Benini et al., 2000), determined at 0.97 Å resolution. Such protein contains a *c*-type heme with a hexacoordinate low-spin iron ion, axially bound to His and Met residues. Since the X-ray protein structure shows several bond and angle violations with respect to the Xplor-NIH library, it was first minimized with the Xplor-NIH Powell minimization routine, to be consistent with the structures calculated through Xplor-NIH.

A set of 2639 upper proton-proton distance restraints randomly selected among those closer than 6 Å were generated by adding 1 Å to the distances measured in the minimized structure. Pseudocontact shifts (271 values), self-orientation residual dipolar couplings (129 values) and cross correlations between Curie and dipolar interactions (129 values) were calculated for all N, HN, H^{α} and C^{α} atoms of the protein backbone, with respect to the iron ion. For calculating the pseudocontact shifts and the residual dipolar couplings, the magnetic susceptibility anisotropy tensor parameters as obtained from experimental NMR data (Banci et al., 2002) were used. The tensor has the z axis perpendicular to the heme plane and the x axis along the pyrrole I–pyrrole III direction; the axial and rhombic anisotropies were set to 2.20 and -0.18×10^{-32} m^3, respectively. To these paramagnetism-based restraints a maximum error of $\pm 10\%$ with gaussian distribution was applied. The tolerance on the input data was set equal to 10% of the experimental restraint, with lower limits of 0.15 ppm for pseudocontact shifts, of 0.10 Hz for residual dipolar couplings, and of 0.20 Hz for cross correlations.

We applied the protocol without using the final correct magnetic susceptibility anisotropies. *Ab initio* calculations of 50 structures were performed with theoretical estimated values for the magnetic susceptibility anisotropies, and the tensor was fitted over the best 5 structures. Figure 2 shows the trend of the magnetic susceptibility anisotropy values, for three different initial values, with cycling structure and anisotropies calculations. The figure shows that the convergence is achieved after few cycles with very good accuracy, thus demonstrating the correctness of the protocol.

The best 20 structures (the lowest energy structures) among the calculated 200 structures have a backbone RMSD to the mean of 0.29 Å (Table 1). The total energy was 283 kJ mol^{-1}, the energy related to NOE, pseudocontact shifts, residual dipolar couplings and cross correlations being 0.04, 0.25, 1.72 and 0.67 kJ mol^{-1}, respectively. Figure 3 reports the total energy and its components during the simulated annealing process. The backbone RMSD of the mean structure to the minimized X-ray structure is 0.35 Å.

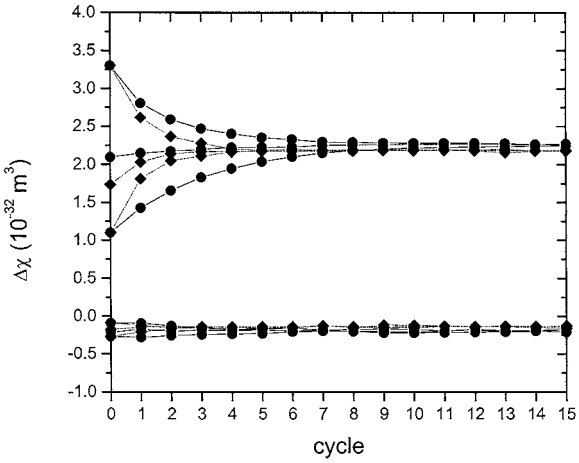


Figure 2. Convergence of the $\Delta\chi$ values ($\Delta\chi_{ax}$: Top lines, $\Delta\chi_{rh}$: Bottom lines) obtained from the fit of the pcs data (●) or the rdc data (♦). Three starting values for the tensor parameters are provided, those obtained with the structure calculated without the paramagnetism-based restraints, and the same increased or decreased of 33%.

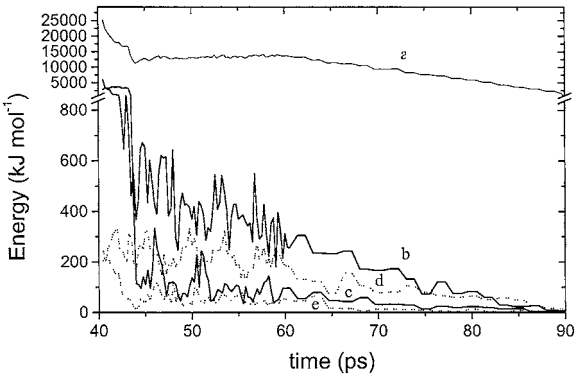


Figure 3. Total energy (a) and its components (NOE: b; pcs: c; rdc: d; ccr: e) during the simulated annealing process for the structure calculation of cytochrome c_{553} using simulated data.

The backbone RMSD of the 20 best structures calculated by including the same NOE restraints only, and excluding all paramagnetism-based restraints, was 0.39 Å, with total and NOE energy of 280 and 0.00 kJ mol^{-1}, respectively (Table 1). Energy values similar to those obtained in the presence of paramagnetism-based restraints indicates that agreement of such restraints does not result in a significant

increase of other energy terms. This is, of course, expected as all restraints are consistent, and proves the efficiency of the paramagnetic package. The backbone RMSD of the mean structure to the minimized X-ray structure is 0.47 Å. This proves that the presence of the paramagnetism-based restraints actually reduces the RMSD and improves the accuracy of the calculated structures.

Structure calculations with experimental data

B. pasteurii *cytochrome* c_{553}

The solution structure of oxidized *B. pasteurii* cytochrome c_{553} was calculated with PARAMAGNETIC DYANA using 1609 meaningful NOEs, 76 dihedral angles and 59 pseudocontact shifts (Banci et al., 2002). Pseudocontact shift values were used as restraints with a tolerance between 0.1 and 0.3 ppm. The program provided values for the axial and rhombic magnetic susceptibility anisotropy of 2.20 ± 0.10 and $-0.18 \pm 0.15 \times 10^{-32}$ m^3, respectively. The BB RMSD to the mean of the family was 0.25 ± 0.07 Å.

The structure was recalculated with the same restraints using Xplor-NIH. The protocol converged to values for the axial and rhombic magnetic susceptibility anisotropy of 1.97 ± 0.09 and $-0.21 \pm 0.16 \times 10^{-32}$ m^3, respectively. The first family, calculated without inclusion of pseudocontact shifts, provided values for $\Delta\chi_{ax}$ and $\Delta\chi_{rh}$ of 1.75 ± 0.20 and $-0.34 \pm 0.19 \times 10^{-32}$ m^3, respectively. The experimental versus calculated values of pseudocontact shifts, for the two cases of such restraints being included or not in the structure calculations, are reported in Figure 4. The tensor is correctly positioned, as shown in Figure 5. The BB RMSD to the mean of the family of the 20 structures with lowest energy on the calculated 200 structures is 0.25 ± 0.04 Å (it is 0.33 ± 0.04 Å for the family obtained without inclusion of the pseudocontact shift restraints). The BB RMSD between the structures obtained with and without including the pseudocontact shift restraints is 0.52 Å. The average energy of the family obtained without including the pseudocontact shift restraints is 288 ± 1 kJ mol^{-1}; its NOE contribution is 0.25 kJ mol^{-1}. The average energy of the family obtained with including the pseudocontact shift restraints is 291 ± 1 kJ mol^{-1}; its NOE and pcs contributions are 0.17 and 3.81 kJ mol^{-1}, respectively. The RMSD between the Xplor-NIH and the DYANA (1K3H) structures is 0.74 Å. The RMSD between the X-ray and the PARAMAGNETIC DYANA structure is 0.61 Å, that between the X-ray and the Xplor-NIH structure is 0.86 Å.

Cytochrome b_5

The solution structure of oxidized rat microsomal cytochrome b_5 is obtained after introduction of 1372 meaningful NOE data, 235 pseudocontact shifts and 62 residual dipolar couplings (Arnesano et al., 1998; Banci et al., 1998a). Two tensors are introduced, one to take into account the paramagnetic susceptibility anisotropy tensor causing pseudocontact shifts and one to take into account the overall molecular magnetic susceptibility tensor causing the residual dipolar couplings, measured from J-modulated experiments at two different magnetic fields. The best 20 structures among 200 calculated structures have a BB RMSD to the mean 0.59 ± 0.10 Å and the resulting paramagnetic axial and rhombic susceptibility anisotropy values are $3.01 \pm 0.24 \times 10^{-32}$ and $-1.40 \pm 0.22 \times 10^{-32}$ m^3, respectively. The molecular axial and rhombic magnetic susceptibility values are $1.88 \pm 0.23 \times 10^{-32}$ and $-0.71 \pm 0.14 \times 10^{-32}$ m^3, respectively. The average energy of the family obtained without including the paramagnetism-based restraints is 439 ± 4 kJ mol^{-1}; its NOE contribution is 4 kJ mol^{-1}. The average energy of the family obtained with including the paramagnetism-based restraints is 455 ± 6 kJ mol^{-1}; its NOE, pcs and rdc contributions are 6.7, 1.26 and 12.1 kJ mol^{-1}, respectively. The family obtained by using PARAMAGNETIC DYANA has a BB RMSD to the mean 0.58 Å, the paramagnetic susceptibility anisotropy tensor parameters are $2.8 \pm 0.1 \times 10^{-32}$ and $-1.1 \pm 0.2 \times 10^{-32}$ m^3 and the molecular susceptibility anisotropy tensor parameters are $2.20 \pm 0.05 \times 10^{-32}$ and $-1.34 \pm 0.04 \times 10^{-32}$ m^3.

Calbindin D_{9k}

The protein calbindin D_{9k} was extensively studied from our group in order to test/apply the use of paramagnetism-based restraints (Allegrozzi et al., 2000; Bertini et al., 2001a, d, 2002a; Barbieri et al., 2002). The protein contains two diamagnetic calcium(II) ions, which can alternatively be substituted with paramagnetic lanthanide(III) ions without alteration of the protein structure.

The complete set of paramagnetism-based restraints is available for this proteins, and they have been included for solution structure calculations with Xplor-NIH. Calculations have been done after introduction of 1611 meaningful NOE data, 105 dihedral angles, 549 pseudocontact shifts, 60 residual

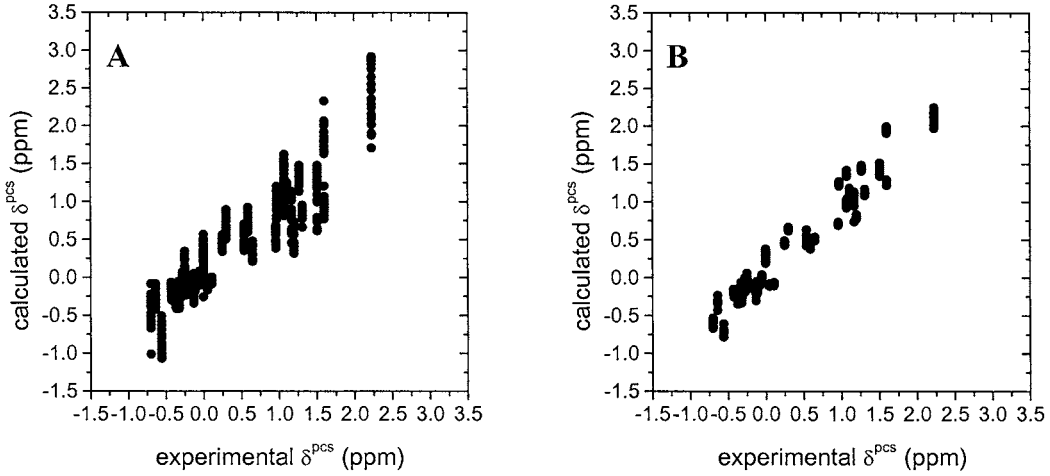


Figure 4. Experimental versus calculated pcs for the cytochrome c_{553} structures obtained without (A) and with (B) the use of the pseudocontact shifts for structure calculation.

dipolar couplings, 26 relaxation rates, 49 cross correlations between Curie and dipolar interactions for the cerium(III) substituted sample; 62 pseudocontact shifts and 20 residual dipolar couplings for the dysprosium(III) substituted sample; and 101 pseudocontact shifts and 37 residual dipolar couplings for the ytterbium(III) substituted sample (Bertini et al., 2001a; Barbieri et al., 2002). Three tensors have been introduced to account for Ce(III), Dy(III) and Yb(III) magnetic susceptibility tensors. Pseudocontact shifts, residual dipolar couplings and cross correlations relative to the same metal are referred to the same tensor. In fact the residual dipolar couplings were experimentally obtained by subtracting the HN ^1J values of the diamagnetic sample from the HN ^1J values of the paramagnetic sample. Therefore, the same anisotropies are introduced in Equations 2 and 4, and the latter are calculated by fitting the pseudocontact shift values, as more accurate than the residual dipolar couplings. The protocol converged to the following tensor anisotropies: $1.97 \pm 0.10 \times 10^{-32}$ and $-0.66 \pm 0.07 \times 10^{-32}$ m^3 for Ce(III) $\Delta\chi_{ax}$ and $\Delta\chi_{rh}$, respectively; $34.1 \pm 1.9 \times 10^{-32}$ and $-21.1 \pm 1.4 \times 10^{-32}$ m^3 for Dy(III) $\Delta\chi_{ax}$ and $\Delta\chi_{rh}$, respectively; $7.46 \pm 0.24 \times 10^{-32}$ and $-3.48 \pm 0.39 \times 10^{-32}$ m^3 for Yb(III) $\Delta\chi_{ax}$ and $\Delta\chi_{rh}$, respectively. These values agree remarkably well with the values obtained using PARAMAGNETIC DYANA (Bertini et al., 2001). The BB RMSD to the mean of the best 20 structures is 0.50 ± 0.08 Å. The average energy of the family is 413 ± 4 kJ mol^{-1}; its NOE, dihedral, pcs, rdc and ccr contributions are 17, 2.5, 12, 10 and 10 kJ mol^{-1}, respectively. The BB RMSD to the mean of the best 20 structures of the family calculated using diamagnetic restraints only is 0.61 ± 0.08 Å; the average energy is 355 ± 1 kJ mol^{-1}; its NOE and dihedral angles contributions are 5 and 1 kJ mol^{-1}, respectively.

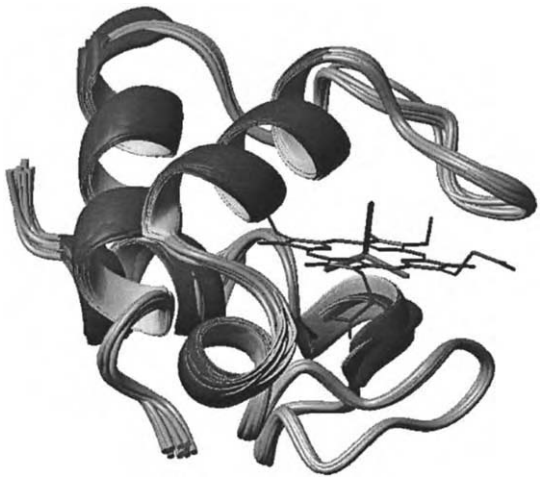


Figure 5. Calculated structure of cytochrome c_{553}. The magnetic susceptibility tensor axes are also shown.

Finally, the module XANGLE has been tested by calculating the solution structure of the protein with the same restraints indicated above but excluding the rdc, and then by providing the polar angles defining the orientation of the NH vectors. These values were obtained by fitting the residual dipolar couplings measured on the Ce(III), Tb(III), Dy(III), Ho(III), Er(III), Tm(III) or Yb(III) substituted protein, as described in (Barbieri et al., 2002). In the absence of the θ and ϕ restraints, the family of the best 20 structures

has a BB RMSD 0.59 ± 0.10 Å, with average energy of 361 ± 1 kJ mol^{-1}, whereas in the presence of the θ and φ restraints, the family of the best 20 structures has a BB RMSD 0.47 ± 0.08 Å, with average energy of 371 ± 3 kJ mol^{-1}.

Concluding remarks

Paramagnetic relaxation enhancements, pseudocontact shifts, residual dipolar couplings due to partial orientation, and cross-correlations between Curie relaxation and nuclear-nuclear dipolar relaxation have been implemented as restraints in Xplor-NIH through dedicated modules and/or protocols. In particular, for paramagnetic relaxation enhancements the Xplor-NIH NOE module is used within protocols for an effective and cautious use of the restraints. The same holds for pseudocontact shifts, for which the use of a tolerance is recommended. A bootstrap Monte Carlo approach is implemented to evaluate the error on the magnetic susceptibility parameters. Such tensor is introduced in a module written by modifying the already available residual dipolar coupling module of Xplor-NIH, in order to efficiently use the metal-based contribution to the alignment of the metalloprotein in high magnetic fields (calbindin case). Alternatively, the residual dipolar couplings due to the overall magnetic anisotropy of the molecule can be used (cytochrome b_5 case). In this case, the overall magnetic anisotropy tensor is obtained, and the resulting values analyzed with the bootstrap Monte Carlo approach. Finally, cross-correlations between Curie relaxation and nuclear-nuclear dipolar relaxation can be provided as restraints after evaluation of a constant which depends on the observed nuclei, on the metal ion, on temperature and on the protein rotational time.

The paramagnetic patch and the file saPARA.inp can be downloaded from the web site:
http://www.postgenomicnmr.net.

Acknowledgements

Marius Clore and Charles Schwieters are acknowledged for providing us the Xplor-NIH code and for their assistance. This work has been supported by the European Commission, contract HPRI-CT-2001-00147, contract HPRN-CT-2000-00092 and contract QLG2-CT-1999-01003.

References

Allegrozzi, M., Bertini, I., Janik, M.B.L., Lee, Y.-M., Liu, G. and Luchinat, C. (2000) *J. Am. Chem. Soc.*, **122**, 4154–4161.

Arnesano, F., Banci, L., Bertini, I., Faraone-Mennella, J., Rosato, A., Barker, P.D. and Fersht, A.R. (1999) *Biochemistry*, **38**, 8657–8670.

Arnesano, F., Banci, L., Bertini, I. and Felli, I.C. (1998) *Biochemistry*, **37**, 173–184.

Banci, L., Bertini, I., Bren, K.L., Cremonini, M.A., Gray, H.B., Luchinat, C. and Turano, P. (1996) *J. Biol. Inorg. Chem.*, **1**, 117–126.

Banci, L., Bertini, I., Bren, K.L., Gray, H.B., Sompornpisut, P. and Turano, P. (1997) *Biochemistry*, **36**, 8992–9001.

Banci, L., Bertini, I., Ciurli, S., Dikiy, A., Dittmer, J., Rosato, A., Sciara, G. and Thompsett, A. (2002) *ChemBioChem*, **3**, 299–310.

Banci, L., Bertini, I., Huber, J.G., Luchinat, C. and Rosato, A. (1998a) *J. Am. Chem. Soc.*, **120**, 12903–12909.

Banci, L., Bertini, I., Cremonini, M.A., Gori Savellini, G., Luchinat, C., Wüthrich, K. and Güntert, P. (1998b) *J. Biomol. NMR*, **12**, 553–557.

Barbieri, R., Bertini, I., Cavallaro, G., Lee, Y.M., Luchinat, C. and Rosato, A. (2002) *J. Am. Chem. Soc.*, **124**, 5581–5587.

Bax, A. and Tjandra, N. (1997) *Nat. Struct. Biol.*, **4**, 254–256.

Benini, S., Rypniewski, W., Wilson, K.S., Van Beeumen, J. and Ciurli, S. (2000) *Biochemistry*, **39**, 13115–13126.

Bentrop, D., Bertini, I., Cremonini, M.A., Forsén, S., Luchinat, C. and Malmendal, A. (1997) *Biochemistry*, **36**, 11605–11618.

Bertini, I., Donaire, A., Jimenez, B., Luchinat, C., Parigi, G., Piccioli, M. and Poggi, L. (2001a) *J. Biomol. NMR*, **21**, 85–98.

Bertini, I., Luchinat, C. and Parigi, G. (2001b) *Solution NMR of Paramagnetic Molecules*, Elsevier, Amsterdam.

Bertini, I., Janik, M.B.L., Liu, G., Luchinat, C. and Rosato, A. (2001c) *J. Magn. Reson.*, **148**, 23–30.

Bertini, I., Janik, M.B.L., Lee, Y.-M., Luchinat, C. and Rosato, A. (2001d) *J. Am. Chem. Soc.*, **123**, 4181–4188.

Bertini, I., Donaire, A., Luchinat, C. and Rosato, A. (1997) *Proteins Struct. Funct. Genet.*, **29**, 348–358.

Bertini, I., Cavallaro, G., Cosenza, M., Kummerle, R., Luchinat, C., Piccioli, M. and Poggi, L. (2002a) *J. Biomol. NMR*, **23**, 115–125.

Bertini, I., Longinetti, M., Luchinat, C., Parigi, G. and Sgheri, L. (2002b) *J. Biomol. NMR*, **22**, 123–136.

Bertini, I., Luchinat, C. and Parigi, G. (2002a) *Progr. NMR Spectrosc.*, **40**, 249–273.

Bertini, I., Luchinat, C. and Parigi, G. (2002b) *Concepts Magn. Reson.*, **14**, 259–286.

Boisbouvier, J., Gans, P., Blackledge, M., Brutscher, B. and Marion, D. (1999) *J. Am. Chem. Soc.*, **121**, 7700–7701.

Brunger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., Read, R.J., Rice, L.M., Simonson, T. and Warren, G.L. (1998) *Acta Crystallogr. D Biol. Crystallogr.*, **54**, 905–921.

Chou, J.J., Li, S. and Bax, A. (2000) *J. Biomol. NMR*, **18**, 217–227.

Clore, G.M. (2000) *Proc. Natl. Acad. Sci. USA*, **97**, 9021–9025.

Clore, G.M. and Bewley, C.A. (2002) *J. Magn. Reson.*, **154**, 329–335.

Clore, G.M. and Garrett, D.S. (1999) *J. Am. Chem. Soc.*, **121**, 9008–9012.

Clore, G.M. and Schwieters, C.D. (2003) *J. Am. Chem. Soc.*, **125**, 2902–2912.

Clore, G.M., Gronenborn, A.M., Brunger, A.T. and Karplus, M. (1985) *J. Mol. Biol.*, **186**, 435–455.

Clore, G.M., Gronenborn, A.M. and Tjandra, N. (1998) *J. Magn. Reson.*, **131**, 159–162.

Clore, G.M., Starich, M.R., Bewley, M., Cai, M. and Kuszewski, J. (1999) *J. Am. Chem. Soc.*, **121**, 6513–6514.
de Alba, E. and Tjandra, N. (2002) *Prog. NMR Spectrosc.*, **40**, 175–197.
Donaldson, L.W., Skrynnikov, N.R., Choy, W.-Y., Muhandiram, D.R., Sarkar, B., Forman-Kay, J.D. and Kay, L.E. (2003) *J. Am. Chem. Soc.*, **123**, 9843–9847.
Dunham, S.U., Turner, C.J. and Lippard, S.J. (1998) *J. Am. Chem. Soc.*, **120**, 5395–5406.
Gaponenko, V., Altieri, A.S., Li, J. and Byrd, R.A. (2002) *J. Biomol. NMR*, **24**, 143–148.
Gochin, M. (2000) *Structure Fold Des.*, **8**, 441–452.
Gochin, M. and Roder, H. (1995) *Protein Sci.*, **4**, 296–305.
Güntert, P. and Wüthrich, K. (1991) *J. Biomol. NMR*, **1**, 447–456.
Güntert, P., Braun, W. and Wüthrich, K. (1991) *J. Mol. Biol.*, **217**, 517–530.
Güntert, P., Mumenthaler, C. and Wüthrich, K. (1997) *J. Mol. Biol.*, **273**, 283–298.
Herrmann, T., Guntert, P. and Wuthrich, K. (2002) *J. Mol. Biol.*, **319**, 209–227.
Huber, J.G., Moulis, J.-M. and Gaillard, J. (1996) *Biochemistry*, **35**, 12705–12711.
Hus, J.C., Marion, D. and Blackledge, M. (2000) *J. Mol. Biol.*, **298**, 927–936.
Kechuan, T. and Gochin, M. (1999) *J. Am. Chem. Soc.*, **121**, 9276–9285.
Kemple, M.D., Ray, B.D., Lipkowitz, K.B., Prendergast, F.G. and Rao, B.D.N. (1988) *J. Am. Chem. Soc.*, **110**, 8275–8287.
Kurland, R.J. and McGarvey, B.R. (1970) *J. Magn. Reson.*, **2**, 286–301.
Moltke, S. and Grzesiek, S. (1999) *J. Biomol. NMR*, **16**, 121–125.
Press, W.H., Flannery, B.P., Teukolsky, S.A. and Vetterling, W.T. (1988) *Numerical Recipes in C–The Art of Scientific Computing*, Cambridge University Press, New York.
Sass, H.J., Musco, G., Stahl, S.J., Wingfield, P.T. and Grzesiek, S. (2001) *J. Biomol. NMR*, **21**, 275–280.
Schwieters, C.D., Kuszewski, J.J., Tjandra, N. and Clore, G.M. (2003) *J. Magn. Reson.*, **160**, 65–73.
Tjandra, N., Garrett, D.S., Gronenborn, A.M., Bax, A. and Clore, G.M. (1997) *Nat. Struct. Biol.*, **4**, 443–449.
Tjandra, N., Marquardt, J. and Clore, G.M. (2000) *J. Magn. Reson.*, **142**, 393–396.
Tjandra, N., Omichinski, J.G., Gronenborn, A.M., Clore, G.M. and Bax, A. (1997) *Nat. Struct. Biol.*, **4**, 732–738.
Tolman, J.R., Flanagan, J.M., Kennedy, M.A. and Prestegard, J.H. (1995) *Proc. Natl. Acad. Sci. USA*, **92**, 9279–9283.
Tolman, J.R., Flanagan, J.M., Kennedy, M.A. and Prestegard, J.H. (1997) *Nat. Struct. Biol.*, **4**, 292–297.
Tu, K. and Gochin, M. (1999) *J. Am. Chem. Soc.*, **121**, 9276–9285.
Turner, D.L., Brennan, L., Chamberlin, S.G., Louro, R.O. and Xavier, A.V. (1998) *Eur. Biophys. J.*, **27**, 367–375.
Werbelow, L.G. (1996) In *Encyclopedia of Nuclear Magnetic Resonance*, Grant, D.M. and Harris, R.K. (Eds.), Wiley, Chichester, pp. 4072–4079.

3.3 Combining in silico tools and NMR data to validate protein-ligand structural models: application to matrix metalloproteinases

[*J. Med. Chem.* in press]

Combining *in silico* tools and NMR data to validate protein-ligand structural models: application to matrix metalloproteinases

Ivano Bertini[1,2]*, *Marco Fragai*[1,3], *Andrea Giachetti*[1,4], *Claudio Luchinat*[1,3], *Massimiliano Maletta*[1,2], *Giacomo Parigi*[1,3] *and Kwon Joo Yeo*[1]

[1]Magnetic Resonance Center, University of Florence, Via Luigi Sacconi 6, 50019 Sesto Fiorentino, Italy

[2]Department of Chemistry, University of Florence, Via della Lastruccia 3, 50019 Sesto Fiorentino, Italy

[3]Department of Agricultural Biotechnology, University of Florence, Via Donizetti 6, 50144 Florence, Italy

[4]ProtEra, Viale delle Idee 22, 50019 Sesto Fiorentino, Italy

* To whom correspondence should be addressed: Tel: +39-055-4574270, Fax: +39-055-4574271, E-mail: ivanobertini@cerm.unifi.it.

ABSTRACT

A combination of *in silico* tools and experimental NMR data is proposed for relatively fast determination of protein-ligand structural models, and demonstrated from known inhibitors of matrix metalloproteinases (MMP). The ^{15}N ^{1}H HSQC spectral assignment and the 3D structure, either X-ray or NMR, are needed. In this method, the HSQC spectrum with or without the ligand is used to determine the interaction region of the ligand. Docking calculations are then performed to obtain a set of structural models. From the latter, the NOEs between the ligand and the protein can be predicted. Guided by these predictions, a number of NOEs can be detected and assigned through a HSQC NOESY experiment. These data are used as structural restraints to reject/refine the initial structural models through further *in silico* work. For a test protein (MMP-12, human macrophage metalloelastase), a final structure of a protein-ligand adduct was obtained which matches well with the full structural determination. A number of structural predictions were then made for adducts of a similar protein (MMP-1, human fibroblast collagenase) with the same and different ligands. The quality of the final results depended on the type and number of experimental NOEs but, in all cases, a well defined ligand conformation in the protein binding site was obtained. This protocol is proposed as a viable alternative to the many approaches described in the literature.

Keywords: ligand-protein docking, NMR guided docking, MMP, flexible docking, NMR restraints

Introduction

Rational drug design strategies must rely on the availability of high-throughput methods to experimentally determine the structure of candidate drug-target complexes.[1] The obtained structural information is then used to improve and optimize the candidate drug in a cyclic procedure. Obtaining three-dimensional macromolecular structures is still a time consuming task. X-ray structure determination is becoming a high-throughput method,[2] but the method requires the easy availability of protein crystals that are suitable for soaking with the various candidate drugs. NMR is also a high-throughput technique in drug discovery,[3,4] but its power lies mostly in the earlier phases of the process, *i.e.* in the first screening of a relatively large number of compounds. NMR quickly provides information on binding affinity and on the region of interaction of the candidate drug with the target molecule.[5]

NMR is of course also able to determine the three-dimensional structure of the adduct, but the procedure is time consuming.[6] Moreover, obtaining a 3D structure depends on the full assignment of thousands of intra-protein NOESY cross peaks, while the only relevant ones are the few intermolecular cross peaks between protein and ligand signals. *In silico* prediction of the structure of the adduct through docking programs, while valuable in the early ligand design phases, is not reliable at this stage.[7-9] Independently of the docking program used, in many cases more than one binding poses are found that do not significantly differ in predicted binding energies.

The availability of a fast and reliable method able to provide a molecular model based on few experimental restraints is an ambitious goal for overcoming these problems. Recently, several efforts have been performed in this direction.[10-13] For instance, a suite of NMR experiments has been recently proposed as a tool to provide structural information on protein-ligand adducts,[12] through intermolecular NOEs detected in selectively labeled proteins. The method is applicable to very large proteins once their three-dimensional structure is known.

For smaller proteins, it is worth to investigate whether a few NOEs may be obtained even without selective labeling of the proteins. We propose here a combined use of computational tools

and a small number of experimental NMR restraints as an efficient way of selecting the correct binding pose among those proposed by docking programs. The experimental restraints are i) the HSQC chemical shifts to select the region of interest on the target, and ii) the few ligand-target NOEs that can be unambiguously identified from ^{15}N NOESY-HSQC experiments. Besides the protein three-dimensional structure, only a singly ^{15}N-labeled protein sample and a pre-existing assignment of its ^{15}N ^{1}H HSQC spectrum are required.

The method has been validated by reproducing the known docked conformation of N-isobutyl-N-[4-methoxyphenylsulfonyl]glycyl hydroxamic acid (NNGH, see chart I) bound to matrix metalloproteinase 12 (MMP-12, human fibroblast metalloelastase). The method has been then applied to obtain the docked conformations of NNGH and other three ligands (3-[[1-[[2-(Hydroxy-methyl)-1-pyrrolidinyl]carbonyl]-2-methylpropyl]carbamoyl]-octanohydroxamic acid (Actinonin), N-[(2R)-2-(hydroxamidocarbonylmethyl)-4-methylpentanoyl]-L-tryptophan methylamide (Galardin), and (2R)-2-mercaptomethyl-4-methylpentanoyl-L-phenylalanyl-L-alanine amide (SIMP-1) (see Chart I) to MMP-1 (human fibroblast collagenase). MMPs belong to a family of zinc-dependent endopeptidases responsible for the metabolism of extracellular matrix proteins,[14-16] and alterations in their levels are implicated in a wide range of pathological states,[17,18] so that these proteins represent attractive drug targets.

Methods

The protocol consists of the following steps, reported in Scheme I: a) identification of the protein binding site, b) calculation of possible protein-ligand adducts, c) prediction of the map of NOEs corresponding to each computed conformation, d) determination of few experimental restraints, able to select the real adduct among those calculated, e) validation and cyclic *in silico* refinement of the ligand position in the protein scaffold. The identification of the protein binding site can be conveniently performed from the analysis of the chemical shifts acquired in the presence and in the absence of the ligand. NOEs between ligand protons and protein protons are obtained

from ^{15}N NOESY-HSQC spectra. The protocol requires that the protein structure and the assignment of its ^{15}N ^{1}H HSQC spectrum is known.

HSQC spectra of the protein in the presence and in the absence of the ligand must be acquired. Most of the protein peaks will coincide in the two spectra. Only peaks corresponding to amide protein protons close to the ligand will be in different positions, but their shift is usually small enough to be easily assigned. This information is used to identify the protein binding site, according to the value of the combined ^{1}H/^{15}N shift perturbation upon complexation, given by $\Delta = (\Delta\delta(^{1}H)^{2}+(\Delta\delta(^{15}N)/6)^{2})^{1/2}$.[19] The residues with a significantly large value of Δ, except those at sizably larger distance from all others, are used to identify the grid for docking calculations. The latter is centered on the protein surface atom closest to the center of the smallest sphere that comprises all the selected nitrogen atoms.

Due to the complexity of the energy landscape on the path to the global minimum region,[20] a specific ligand-protein docking program is invoked in order to accurately probe and select the conformations of the ligand according to appropriate scoring functions. We use the program Autodock because it has been amply validated and tested on the target proteins selected for this study. The docking program can be run to obtain clusters of the possible adducts. Such clusters are then used to predict NOEs between protein and ligand nuclei. In fact, a map of distances between ligand and protein nuclei can be obtained for each of the different clusters. The presence of cross peaks can thus be predicted for the different possible adducts and compared with the cross peaks actually present in the experimental spectra.

The following experiments must be performed: ^{15}N NOESY-HSQC spectra of the protein-ligand adduct and of the free protein, and the 1D ^{1}H spectrum of the free ligand in water. The latter experiment provides an estimate of where the chemical shifts of ligand signals in the adduct have to be looked for. The presence of intermolecular cross peaks, *i.e.* peaks between frequencies close to those of the free ligand in one dimension, and those of the protein amide protons predicted to be in the vicinity of the ligand in the other dimension, is checked. Such cross peaks, if absent in the free

protein spectrum and not attributable to nuclei of other neighboring protein residues, are unambiguously assigned. A good correspondence between expected and observed cross peaks is a clear indication of the goodness of the corresponding cluster. On the other hand, direct evidence of the non acceptability of some of the clusters generated by Autodock can be obtained. The experimental NOEs, translated into upper distance limits, can then be used to refine the remaining acceptable structures and possibly to further discriminate among them. The refinement procedure has been developed using Xplor-NIH. In such procedure, the protein side chains are left free to move, thus allowing a better docking to be obtained with respect to docking programs where the protein is completely rigid.

The refinement procedure consists in loading the calculated adduct and performing an *in vacuo* molecular dynamics simulation in internal coordinates, with backbone atoms grouped together to constitute a rigid structure. A simulated annealing is performed by heating the system to 1500 K and then cooling it to 50 K in steps of 50 K. At each temperature, 750 steps of molecular dynamics simulations are performed with time steps of 2 fs. The force constant of NOE restraints is fixed to 30 kcal mol^{-1} $Å^{-2}$, and van der Waals, electrostatic terms and the protein and ligand force field (angles, bonds, dihedrals and impropers) are also included. The resulting structures are then refined with a Powell minimization, and ordered according to the value of the target function. The latter is calculated considering the ligand-residue and residue-residue interactions only for residues up to 8 Å from the ligand. This helps reducing the energy “noise” originating from slight changes in residue-residue interactions far away from the ligand site. In all cases, the best 10 structures over 200 calculated through Xplor-NIH starting from each tentative docking structure are very similar to one another.

The structure of the adduct is thus calculated through the consecutive use of the program Autodock and the refinement procedure working in Xplor-NIH. Xplor-NIH calculations can significantly change the protein side chain positions after complexation. Therefore, cycling between Autodock and Xplor-NIH refinement is necessary until convergence to a fixed protein structure is

achieved. We have tested that such approach can actually select the correct ligand-protein docking, among those proposed by Autodock. Furthermore, the introduction of experimental data and the allowed mobility of the protein side chains provide more confidence in the obtained adduct.

MMP systems, the receptors that we used in this work, have a catalytic zinc ion as active center, coordinated to three histidines. The three zinc-coordinated histidines were treated as the neutral form with the hydrogen on ND1, whereas other histidines used the default option with hydrogen on NE2. Glutamates were treated as charged form as default, except the catalytically essential glutamate 219,[21] at the second shell of the zinc binding site. The latter residue was protonated, with the hydrogen on the oxygen nearest to the catalytic zinc, or deprotonated depending on whether the zinc donor atom closest to it was deprotonated (hydroxamate ligands)[21] or protonated (thiol ligands). In order to take into account the electron density delocalization due to coordination of ligands, the charge of the zinc ion was distributed among the protein ligands.[21]

Results

Test with a known structure: MMP-12-NNGH

NNGH is a broad spectrum MMP inhibitor able to interact with both the catalytic zinc and the S1' cavity.[6,22] In particular, it is able to bind MMP-12 with nanomolar affinity (K_d = 10 nM)[6], and for this reason it has been chosen as a model system to study protein-inhibitor interactions. Its molecular structure is reported in Chart I.

The structure of MMP-12 complexed to NNGH is already known,[6] as both the X-ray (Figure 1A) and the NMR structures of the adduct have been solved. Therefore, we used such system as a test for our protocol. The HSQC spectra of the protein without and with the NNGH in solution were acquired. Inspection of residues showing significant chemical shift perturbation (see Table 1) permitted to define the ligand binding region as the protein catalytic site (defined here as constituted by the zinc binding region, the S1' pocket and the substrate binding groove) with reasonable accuracy. As expected from the crystal structure of the MMP-12-NNGH adduct, among the affected

resonances are residues 210, 211, 215 and 216 on the alpha helix at the bottom of zinc binding site, residues 237, 239-240 and 242 forming the hydrophobic S1' cavity, and residues 179-182 and 184 on the strand facing both the catalytic metal and the S1' pocket (Figure 1B).

Calculations were performed using the X-ray structure of the protein (PDB 1Y93) at 1.03 Å resolution.[6] Autodock was used to select the lower docking energy conformations. Docked conformations were clustered according to a maximal RMSD of 1 Å (Figure 2). The docking energies for the first, second, third and fourth clusters were –15.89, –15.84, –15.01 and –14.44 kcal mol^{-1}, respectively. The second cluster is in accordance with the X-ray structure of the adduct (PDB 1RMZ).[6] The plane containing the hydroxamic group in the first and third cluster is oriented perpendicularly to the plane containing the hydroxamic group in the second cluster. The *p*-methoxy-phenyl group enters more deeply in the S1' pocket in the first than in the third cluster. In the fourth cluster the *p*-methoxy-phenyl group does not sit in the S1' pocket.

These structures were separately refined with Xplor-NIH using the already available NOEs with protein backbone NH atoms (see Figure 3B).[6] The second cluster remains essentially unchanged, with total energy –1184 kcal mol^{-1} (see Figure 2). The structure calculated using the third cluster as starting conformation is similar to the previous one, with total energy –1179 kcal mol^{-1}. The structure calculated from the first cluster has total energy –1062 kcal mol^{-1}, and no coordination of the hydroxamic group to the metal ion; the one calculated from the fourth cluster has total energy –734 kcal mol^{-1}, and the *p*-methoxy-phenyl group outside the S1' pocket.

Slight changes in the side-chain protein structure were obtained, and new Autodock calculations were thus performed using the three lowest energy Xplor-NIH protein structures. Remarkably, the lowest docking energy clusters calculated by Autodock now converge to similar conformations using the second and third Xplor-NIH protein structure (see Figure 2). These conformations are in agreement with the X-ray structure, with docking energy from –15.85 kcal mol^{-1} to –15.63 kcal mol^{-1}. Xplor-NIH refinements provided structures (see Fig. 2 and Fig. 3B)

with lowest total energy from −1197 to −1180 kcal mol^{-1}, in agreement with the X-ray structure (see Figure 3A).

Analogous calculations were performed also using the X-ray structure PDB 1OS9, with 1.85 Å resolution.[23] In this structure the active site of one molecule is not hosting an external ligand but the N-terminal part of the neighboring protein molecule. The calculations converged to the same adduct obtained starting from the 1Y93 structure.

Determination of structural models for ligand adducts of MMP-1

NNGH itself and three other known strong inhibitors of MMPs were selected as representatives of different classes of ligands and tested against MMP-1. The test consists in following the protocol described above and checking whether i) unambiguously NOEs could be obtained and ii) the cycling between Autodock and Xplor-NIH calculations permits the selection of one ligand conformation. Calculations were performed using the X-ray structure of the inhibitor-free protein (PDB 1CGE) with 1.90 Å resolution.[24]

MMP-1-NNGH

The first ligand examined is the same ligand used to validate the protocol with MMP-12. The structure of the NNGH adduct with MMP-1 is not known, although it is reasonable to believe that it will adopt a similar conformation. We measured an IC_{50} value for the adduct of 174 nM.

Chemical shift perturbation affects the zinc binding histidine 228 and the neighbouring residues 226, 227 and 229, residues 239-240 and 243 forming the S1' hydrophobic pocket, residues 215, 217 and 219 on the alpha-helix where the metal binding site is inserted, and residues 180 and 184 on the parallel strand (see Fig. 1C). This is an expected feature, but it is a new independent experimental information based on which an Autodock grid was generated. The grid resulted nicely centered around the known catalytic site. Autodock calculations using this grid were thus performed.

The four lower docking energy clusters were analysed. The docking energies were –14.68, –13.99, –13.94, –13.81 kcal mol^{-1}, respectively. The structures in the first and third clusters show similar hydroxamate coordination to the catalytic zinc. The ligands in the second cluster are oriented similarly to those in the first cluster, but the hydroxamic acid is coordinated to zinc only though the carboxylic oxygen. The structures in the fourth cluster show coordination of the sulphonate oxygen (SO) atoms to zinc. In all cases the *p*-methoxy-phenyl group sits in the S1' hydrophobic pocket. The position of the *i*-butyl group changes in the four adducts. In the first and second clusters it prevents the formation of hydrogen bonding between the hydroxamic HN and alanine 182 oxygen, whereas the latter hydrogen bond is present in the third cluster.

NOE restraints were obtained in the following way. In the ^{15}N NOESY-HSQC spectrum, a cross peak at chemical shift of 11.7 ppm is present in the N leucine181 plane (Fig. 4). Such a shift is too high to be assigned to a protein signal, as no tryptophan residue is close to the active site. Therefore, it was assigned to the unique amide proton of NNGH. Aromatic protons of NNGH are close to N of residues glycine 221, histidine 222, alanine 216 and arginine 214 according to the structures calculated by Autodock. We have searched in the spectrum all the long range NOEs between aromatic protons and amide groups of these residues. New peaks in the spectrum of the adduct, which cannot be due to intraprotein interactions, actually appear in the aromatic region (Fig. 4), and were assigned as reported in Table 2.

The structural families obtained with Xplor-NIH starting from the first three lowest Autodock docking energy structures converged to the same conformation (see Fig. 5). This conformation was similar to the conformation of the third Autodock cluster, with the exception that the sulfur oxygen H-bonded to alanine 182 was the most external oxygen atom rather than the internal one. The lowest total energies were –622, –617, –615 kcal mol^{-1}, respectively. The lowest total energy of the structural family obtained with Xplor-NIH starting from the fourth Autodock structure was –592 kcal mol^{-1}. This adduct, slightly different from the other three for the fact that zinc coordination by hydroxamate was loose, can be excluded due to its larger energy.

No appreciable changes in the protein sidechain positions are observed and thus further Autodock/Xplor-NIH cycles were not needed. Therefore, the structural family shown in Fig. 3C represents an experimentally validated and unique structural model for the MMP-1-NNGH adduct.

MMP-1-Actinonin

Actinonin, whose molecular structure is reported in Chart I, is a well known inhibitor of aminopeptidases and peptide deformylase.[25] It is also a strong inhibitor for some MMPs, with a K_i of 300 nM for its adduct with MMP-1.[26]

Chemical shift perturbations again allow us to map the region of interest on the protein surface. Residues 215, 217, 218, 220, 223 and 227 forming the metal binding site, residues 235, 236 and 249 on the loop that covers the S1' pocket, and residues 180 and 182 on the spatially close strand (see Figure 1D) define the ligand binding region, and were used for the definition of the Autodock grid. Despite the non complete correspondence of the affected residues with those found for the NNGH adduct, the resulting grid was quite similar. Four clusters were then calculated (docked conformations were again clustered according to a maximal RMSD of 1.0 Å, see Figure 6). In all structures the hydroxamate is bound to the catalytic zinc. However, whereas in the first two structures the pentyl group is located inside the S1' hydrophobic pocket and the external propyl group is differently oriented, in the third and fourth structures the two groups are interchanged. The lowest docking energy for the structures in the four clusters were –19.91, –19.11, –18.82, –18.62 kcal mol^{-1}, respectively.

Cross peaks of all protons belonging to the ligand with the HN protein protons expected at distances shorter than 5 Å for one or another cluster were looked for in the ^{15}N NOESY-HSQC spectrum. Since the NH of tyrosine 240 has two unassigned cross peaks at frequencies typical of methyls, they must be related to two methyl groups which are close in the structure of the adduct. From the clusters generated by Autodock, they can only be H' and H'''. The following peaks were thus assigned: (a) methyl protons H''' with tyrosine 240 and with the aligned threonine 241; and (b)

H' with tyrosine 240. Among the clusters generated by Autodock, the third and fourth clusters can be readily excluded, because in such structures the above cross peaks could not be observed. Therefore, by looking at the other two clusters, we also assigned the following cross peaks, which cannot be assigned to other intraresidue protons or to sidechain protons of close residues: (c) alanine 184 with H1, and tyrosine 240 with H6, as such protons are the closest to the coupled HN protons; (d) leucine 181 and tyrosine 240 with H', as they are aligned and close to one another.

Xplor-NIH calculations were thus performed to refine the selected Autodock structures. Actually, we performed the calculations not only starting from the first two structures, but also starting from the structures excluded according to the observation of the ^{15}N NOESY-HSQC spectrum. The first and second family of structures calculated with Xplor-NIH are very similar to the corresponding Autodock structures; the third Xplor-NIH structural family shows significant rearrangements in the position of the ligand branches, but the pentyl group remains located outside the hydrophobic pocket; in the fourth Xplor-NIH structural family the pentyl group lies in the hydrophobic pocket, thus resulting similar to the first and second families. Xplor-NIH energies for the four families are –870, –862, –811 and –840 kcal mol^{-1}, respectively. This indicates that the third structure, quite different from the other three, is not acceptable. The calculations show that the method is indeed robust. In fact, the first two Autodock structures that were selected from the observation of the NMR spectra actually have the lowest energy, whereas the third has a sizably larger energy even after Xplor-NIH refinement. Interestingly, the fourth Autodock structure, initially completely different from the first two, was brought by Xplor-NIH calculations to converge with the first two.

A second Autodock and Xplor-NIH cycle was performed starting from the lowest energy protein structure. The calculated Xplor-NIH structures, in fact, showed slightly different positions of protein sidechains, in particular of residues leucine 181, proline 238 and tyrosine 240. Such new protein conformation was provided to Autodock for a new docking calculation. The best four Autodock clusters (docking energy –20.61, –19.07, –18.99, –18.88 kcal mol^{-1}) were then provided

to Xplor-NIH. The first, third and fourth clusters display both the hydroxamate and the pentyl group in similar positions; the second one is completely different (the hydroxamate does not bind the zinc ion). All lowest energy Xplor-NIH structures (see Fig. 3D), with the exception of those calculated starting from the second Autodock structures, converged to the third Autodock conformation, and are equivalent to the lowest energy Xplor-NIH family calculated in the first cycle. The total energies for these structures are –881, –880 and –870 kcal mol^{-1}. The Xplor-NIH structure calculated starting from the second Autodock structure has a total energy of –810 kcal mol^{-1}, and can thus be excluded. Therefore, the structure family of Fig. 3D is a unique structural model for the MMP-1-Actinonin adduct.

MMP-1-Galardin

Galardin (see Chart I) is a broad spectrum peptidomimetic inhibitor of MMPs[16] with an IC_{50} of 1.5 nM for MMP-1.[27] Chemical shift perturbation involved residues 215, 216, 218, 220, 223 and 227-228 at the metal binding site, 236-237, 240 and 243 at the large loop covering the S1' cavity, and 179 and 183 at the strand facing the S1' cavity and the metal binding site (Figure 1E). These residues were used to define the Autodock grid, which again was found very similar to the previous ones. The four lowest docking energy clusters calculated by Autodock (–19.79, –19.51, –18.59, –17.48 kcal mol^{-1}, respectively) showed the following features (see Figure 7). In the first, second and third cluster the *i*-butyl group enters the S1' pocket, whereas in the fourth cluster it is outside. The structures in the first and second clusters are very similar, as they differ only for the orientation of the indole group, positioned outside the S1' pocket. The structures in the third and fourth cluster are quite different from those in the first and second cluster, including the position of the indole group, which in any case remains outside the S1' pocket.

In the ^{15}N NOESY-HSQC spectrum, two cross peaks are present in the N phenylalanine 242 and tyrosine 240 planes. Such peaks are at chemical shifts typical of methyl groups and cannot be assigned to any intraresidue proton or proton of close residues. Since in Galardin there are three

methyl groups, two of them being close in the structure, the latter (H' and H'') were assigned to these peaks. Another cross peak is present, which cannot be assigned to protein protons, in the plane of tyrosine 240. This cross peak falls into the aliphatic region, and therefore it could be provided by CH or CH_2 protons. Since such proton must be close to H' and H'', which have also a cross peak with tyrosine 240, it was assigned to H6 or H7.

Xplor-NIH calculations change only slightly the conformations obtained with Autodock relatively to the first three clusters. The structure obtained starting from the fourth Autodock cluster is instead modified by the NOE restraints to have the *i*-butyl group inside the S1' pocket as in the other three clusters. The total energy of the Xplor-NIH structures are –575, –578, –565 and –527 kcal mol^{-1}, respectively. Only small changes in the side chain positions have been observed, regarding in particular residues from 238 to 241.

The three protein structures with the smallest Xplor-NIH energy were used to repeat Autodock calculations. In the first case Autodock produced the two lowest docking energy clusters very similar to those obtained in the first run (–18.87 and –18.57 kcal mol^{-1}), whereas the third and fourth clusters (with docking energy –18.08 kcal mol^{-1}) have now the indole group inside the S1' pocket. These conformations can be excluded by the observed NOEs. It is remarkable that such faulty Autodock behaviour occurs in the second round, i.e. after adjustment of the structure by Xplor-NIH minimization. This observation underlines the need for experimental restraints to gain confidence in *in silico* models. In the second case, the three lowest docking energy clusters are again very similar to those obtained in the first run (–19.59, –19.56 and –18.14 kcal mol^{-1}), whereas in the fourth cluster (with docking energy –17.83 kcal mol^{-1}) the *i*-butyl group is outside the S1' pocket. In the third case, the lowest docking energy cluster is again similar, with energy –18.42 kcal mol^{-1}. The Xplor-NIH calculations performed with the four lowest docking energy structures as starting conformations converged to a unique conformation (–586, –584, –576 and –576 kcal mol^{-1}), except for the indole group which, being outside the S1' pocket, is free to move (Figure 3E).

Again, the family of Fig. 3E can be confidently taken as a validated structural model for the Galardin adduct of MMP-1.

MMP-1-SIMP-1

SIMP-1 is a polypeptide derivative able to inhibit collagenases.[16] Its molecular structure is reported in Chart I. We measured an IC_{50} value for the adduct of 46 nM. In the MMP-1-SIMP-1 adduct, affected resonances include residues 215, 217-218, 220, 222-223 and 227-228 on the alpha helix of the zinc binding site, residues 235-237, 239-240 and 242 forming the S1' cavity, and residues 179-180 and 184 on the strand facing both the catalytic metal and the S1' pocket (Figure 1F). The four clusters with smallest docking energy calculated by Autodock (see Figure 8) have docking energy of –16.15, –15.92, –15.76 and –15.72 kcal mol^{-1}. In the first and second clusters the sulfur atom coordinates the catalytic zinc; in the first cluster the S1' pocket interacts with the ligand benzyl group, in the second with the *i*-butyl group. In the third cluster the sulfur atom is hydrogen bonded to the oxygen of glycine 179, on the other site of the catalytic pocket with respect to the zinc ion, and the ligand benzyl group sits in the S1' pocket. In the fourth cluster, the ligand is oriented similarly as in the first cluster, but the ligand sulfur atom is loosely coordinated to the zinc ion, and hydrogen bonded to glutamate 219.

In the ^{15}N NOESY-HSQC spectrum, in the N plane of residue leucine 181, there are two signals in the aromatic region which cannot be assigned to protein side chains. Therefore, they must be assigned to protons of the aromatic ring of the SIMP-1. Two cross peaks, one of low and one of high intensity, are present in the N plane of tyrosine 240 at frequencies typical of methyl groups, which cannot be assigned to intraresidue or sequential contacts. In one of the clusters calculated by Autodock, N of tyrosine 240 is close to two of the three methyls of SIMP-1, H' being closest than H'', and thus the cross peaks were correspondingly assigned (see Table 2). A further distance restraint is determined from another cross peak in the N plane of threonine 241, aligned with the signal assigned to H''.

Xplor-NIH calculations select the second Autodock cluster as the correct one. In fact it remains almost unchanged after refinement, with total energy –709 kcal mol^{-1}. Calculations performed starting from the other clusters provide structures very different from the starting ligand conformation, and with the ligand not coordinated to the zinc ion. Their total energies are larger than –634 kcal mol^{-1} and such structures are thus excluded.

Slight changes in the protein side chain positions are observed, in particular on residues 180, 214 and 219. A second Autodock calculation was thus performed. The first three clusters (with docking energy of –16.15, –16.12 and –15.88 kcal mol^{-1}, respectively) show a ligand pose similar to that calculated in the first and fourth clusters of the first Autodock run. The fourth cluster, with docking energy –15.77 kcal mol^{-1}, is instead similar to the pose already identified as correct. Xplor-NIH calculations again confirmed such structure as the correct one, with total energy –708 kcal mol^{-1}. The corresponding family is shown in Fig. 3F. This family represents the validated structural model of the MMP-1-SIMP-1 adduct.

Backbone mobility

In order to test the protocol for possible protein backbone rearrangements upon complexation, Xplor-NIH calculations were also performed with allowing the protein backbone to move in the protein region affected by chemical shift perturbation. In all cases we found no appreciable differences in the results. In fact, for all adducts the lowest energy structures corresponded to those identified as correct in the calculations performed with rigid backbones.

Discussion

A protocol has been developed to merge the “pure” docking capability of Autodock (or other docking programs) with the exploitation of available experimental restraints. For the relatively strong ligands (K_{diss} ~ μM or less) elected here, the protocol has been shown to be efficient, robust and reliable. As shown in Scheme I, the protein binding site is identified from chemical shift

perturbation in the HSQC spectrum of the protein upon complexation. The observation of shift perturbations on passing from the assigned spectrum of the free protein to the spectrum of the adduct permits the definition of the protein grid to be used in Autodock calculations. Autodock usually provides several clusters of structures for the adduct, which often have similar docking energy. These structures are used to calculate maps of NOEs, to be compared with NOEs actually observed in the ^{15}N NOESY-HSQC spectrum of the adduct. A few ligand-protein NOEs can always be assigned, and the latter can be used as restraints in Xplor-NIH calculations for selection, validation and refinement of the Autodock structures. One-two cycles at most may be needed in case Xplor-NIH calculations modify some protein side chain positions with respect to the structure provided to Autodock. All these steps could be performed semi-automatically, if required.

The protocol relies on the following information to be available: the protein structure; the assigned HSQC spectrum of the free protein; the ^{15}N NOESY-HSQC spectrum of the free protein; the HSQC and ^{15}N NOESY-HSQC spectra of the protein-ligand adduct; and the 1D ^{1}H spectrum of the ligand. The protocol has been developed in order to avoid preparation of doubly-labelled samples and assignment of protein side chains, thus resulting in a much faster throughput.

We have shown that such approach is actually efficient in finding the protein-ligand structure for four adducts of MMP-1 with different ligands. The peculiarity which makes this approach successful in the cases here examined is the combination of a docking program, able to quickly and efficiently sample the possible binding poses, with a molecular dynamics program, which selects the proposed poses using few non ambiguous experimental data. In this way the efficiency of the former program is coupled to the complexity of the latter, which also allows for protein side chain movements. The program has been deliberately tested using only non ambiguous NOEs obtainable from the assignment of HN, but it is obviously open to the use of additional or different restraints. We decided to use the chemical shift perturbations only for the determination of the grid to be used for the docking program calculations, without including them as restraints in the molecular dynamics program due to their ambiguous nature, although ambiguous restrains could be

in principle used, either as such, as recently proposed,[13] or through calculation of *j*-surfaces.[10] The use of chemical shift perturbations for the determination of the grid is much less stringent than their use as constraints, as a few "second sphere" shifts erroneously mistaken for first sphere shifts may drive the ligand in wrong positions, while the resulting grids are expected to be only somewhat broadened. As a matter of fact, differences in perturbed residues from one ligand to another do not result in grossly different grids, and the latter, in all cases, encompassed the whole catalytic site.

Several predicting programs for protein-ligand adducts have been proposed in the literature. Inclusion of biochemical and biophysical data in docking protocols, called guided docking,[28,29] is a common approach to reduce the conformational variety of the proposed solutions. Some other programs[7-9,30-35] work totally *in silico*, without experimental information on the investigated adduct, and perform docking calculations with an improved level of sophistication. They can be successful, but the level of confidence for the proposed adduct is difficult to establish. Furthermore, a strong bias towards known solutions or preconceived requirements is introduced if the docking is restrained according to chemical information derived from databases of protein-ligand complexes. Other programs[36-39] use the experimental NMR information more systematically, thus being similar to structural determination programs and therefore more time consuming. NMR-derived restraints were also used in docking programs to identify the location of the ligand binding[10] and to restrict the conformational space for molecular modeling routines.[11] NMR experiments on selectively labeled proteins were also used to obtain structural information on protein-ligand complexes.[12] This approach, although more expensive than the one here proposed, is probably the only viable in case of large proteins. To our knowledge this is the first time that an approach is proposed where few experimental data are used to select and refine poses proposed by fast docking programs.

Autodock has been selected among the docking programs because in the case of MMPs it was demonstrated to be a robust program with good docking accuracy and reliability, including the correct geometry of the zinc binding groups.[21,40] It employs a genetic algorithm searching function, able to efficiently sample large search spaces. Different docking programs could however be used if

considered more reliable in other cases. In the same way, other molecular dynamic programs could be used instead of Xplor-NIH. We used Xplor-NIH as an NMR-oriented wide spread general program for structural calculations using simulated annealing. Ligand growing procedures[30] may also be implemented in Xplor-NIH, resulting probably useful especially in case of large ligands.

Although the presence of the metal ion in MMPs tends to restrict the number of Autodock clusters by favoring poses where the hydroxamic moiety is coordinated to the metal, the protocol is expected to be useful also in case of proteins not containing catalytic ions. Actually, docking programs are developed to work mainly in their absence, and, in case they propose several different conformations, the detection of NOEs may result decisive for the selection of the correct one. Indeed, as we have seen, Autodock does not always succeed in correctly binding the metal to the hydroxamic moiety. Furthermore, in the absence of the metal, further H-bonds or van der Waals contacts should occur for strong ligands, which would likely provide additional intermolecular NOEs.

We have shown that it is possible to obtain few intermolecular experimental NOEs through fast NMR experiments without the necessity to assign all protein NOESY cross peaks. Only unambiguous NOEs between protein and ligand protons have been considered; therefore, cross peaks were assigned to ligand protons only if they could not be reasonably assigned to any protein side chain proton, taking into account the structural adducts proposed by Autodock. In all cases here addressed, experimental restraints have been shown to be necessary and sufficient to extract the adduct conformation among the several proposed by Autodock with similar docking energy, and thus are used to validate them. Furthermore, the approach proposed can also be useful to refine the structure of the ligand-protein adduct, especially because local small modifications in the protein structure (of side chains, if sufficient as in the present case, but also in the protein backbone, if needed – see below) can be accommodated by cycling between Autodock/Xplor-NIH runs. This makes the present approach preferable to the direct introduction of distance restraints in docking programs with a fixed protein matrix.

The solution structure of the inhibitor-free MMP-1, obtained from a series of 3D triple-resonance NMR experiments, shows nearly identical both backbone and secondary structures than the crystallographic structures.[41] Furthermore, the backbones of the solution structures of the inhibitor-free MMP-1 and of the MMP-1 complexed with a sulfonamide derivative of the hydroxamic acid compound have been shown to be essentially identical,[42] although mobility measurements indicate that the region near the active site is highly mobile.[41,42] It is thus reasonable, at least in our case, to assume that the protein backbone remains rigid during complexation in solution, and with structure identical to the crystallographic structure.

Although not necessary for the present calculations, also the protein backbone could be allowed to (partially) move in Xplor-NIH calculations (see results). This could be important if modest backbone rearrangements are expected upon ligand binding, as could be indicated by chemical shift perturbations spread out over a wider region.

It is known that effective MMP inhibitors achieve tight binding via extensive van der Waals contacts with the hydrophobic interior of S1' and by strong electrostatic interactions with zinc and nearby charged or polar side chains.[43] All calculated adducts indeed show ligand coordination to the catalytic zinc and the formation of a net of hydrogen bonds between ligand and protein residues. This result is not trivial as it may seem, as several of the initially obtained Autodock structures had severely distorted – or were even lacking – hydroxamate coordination to the zinc ion.

The distance between zinc and hydroxamate oxygens is in all calculated structures between 1.95 and 2.25 Å. The O-Zn-O angle is always between 86 and 93°. The coordination geometry is distorted square-pyramidal in MMP-12-NNGH and MMP-1-Actinonin, and distorted trigonal bipyramidal, with hydroxamic O2 and N histidine 222 in axial positions, in MMP-1-NNGH and Galardin. All hydrogen bonding interactions between MMPs and ligands are reported in Table 3. In particular, H-bonds are present in all adducts with NNGH, Actinonin and Galardin between oxygen of alanine 182 and the amide proton of the hydroxamic group, as well as between the protonated glutamate 219 and the oxygen of the hydroxamic group. H bonds are also present between ligands

and HN of Leu 181, as previously seen in the MMP-1-CGS[42] and in the MMP-12-NNGH adducts. In the MMP-1-SIMP-1 adduct, with a distorted tetrahedral coordination geometry around the zinc ion, constituted by the three histidine nitrogen atoms and the sulfur SIMP-1 atom, a net of hydrogen bonds is formed, connecting the ligand to the protein atoms (see Table 3). Both the coordination geometry and the H-bonding network can be used to assess the reliability of the obtained adducts. In all the adducts, the inhibitors establish enough interactions to reach nanomolar affinity. In particular, all the ligands bind the metal, place a lipophylic moiety into the S1' cavity and establish two or more hydrogen bonds with atoms of the protein groove. This binding mode is reasonable and is indeed adopted by many strong ligands of MMPs.

Cycling between fast docking programs and Xplor-NIH calculations can be used to assess ligand-protein structures also in the presence of restraints different from NOEs. Diamagnetic residual dipolar couplings have already been demonstrated to be extremely useful to predict the structure of protein-protein adducts.[20,44-46] Also pseudocontact shifts have been used for the study of protein-protein docking.[47] Paramagnetism-based restraints, and in particular paramagnetic relaxation rates, pseudocontact shifts and residual dipolar couplings, arising when a paramagnetic metal ion is coordinated to the protein, could be employed as restraints in the proposed protocol for protein-ligand docking. Xplor-NIH has the advantage that it already contains the tools needed to deal with such restraints.[48]

Conclusions

A novel protocol to obtain validated structural models of protein-ligand complexes has been developed and applied for the determination of the structure of the adducts of the protein MMP-1 with four different ligands. The method was shown to be reliable, as tested for the known structure of the adduct of one of these ligands with MMP-12. It uses NMR derived restraints obtained using singly (^{15}N) labelled proteins. The strategy that we propose promises to be generally useful also for

the structural determination of different protein-ligand adducts, whenever the structure of the free protein is known and the structural changes upon complexation are not expected to be dramatic.

Experimental section

Sample preparation

The fragment of human fibroblast collagenase corresponding to proMMP-1 (Pro21-Pro269) and bearing an additional methionine at the N-terminal, was expressed in E. coli. The cDNA was cloned into the pET21 vector (Novagen) using *Nde I* and *Xho I* as restriction enzymes. The *E. coli* strain BL21 Codon Plus cells, transfected with the above vector, were grown in 2 × YT media at 37°C. The protein expression was induced during the exponential growth phase with 0.5 mM of IPTG. Cells were harvested for 4 h after induction. Uniform ^{15}N-labeled protein was obtained by growing the transfected BL21 Codon Plus cells in minimal media at 37°C. The cells were lysed by sonication and the inclusion bodies, containing the proMMP-1, were solubilized in 2 M urea; 20 mM Tris pH 8.0. The protein was purified on the Hitrap Q column (Pharmacia) with a buffer containing 2 M urea and 20 mM Tris (pH 8.0). The elution was performed using a linear gradient of NaCl up to 0.35 M. The purified protein was then refolded by using a multi-step dialysis against solutions containing 50 mM Tris (pH 7.2); 10 mM $CaCl_2$; 0.1 mM $ZnCl_2$; 0.3 M NaCl. The refolded protein was exchanged, by dialysis, against a buffer with 10 mM Tris pH 7.2; 5 mM $CaCl_2$; 0.1 mM $ZnCl_2$; 0.3 M NaCl. The protein was activated by 1 mM APMA (4-aminophenylmercuric acetate) at 4°C overnight and dialyzed with a buffer containing 10 mM Tris pH 7.2; 5 mM $CaCl_2$; 0.1 mM $ZnCl_2$; 0.3 M NaCl; 0.2 M Acetohydroxamic acid (AHA). The activated protein (Val 101- Pro 269) was concentrated using an Amicon stirrer and Centriprep concentrators, fitted with a YM10 membrane in nitrogen atmosphere at 4°C. Catalytic domain of MMP-1 was purified using size-exclusion chromatography with the final dialysis buffer and concentrated up to 0.5 mM using an Centriprep concentrators in nitrogen atmosphere at 4°C. The final protein sample was dialysed

against a solution containing 50 mM sodium acetate, 100 mM NaCl, 5 mM $CaCl_2$, 0.1 mM $ZnCl_2$, with 10% of D_2O (pH 6.5).

Inhibited proteins were prepared by titration of the free-MMP-1 with equimolar amounts of NNGH, SIMP-1, Galardin and Actinonin.

NNGH, Galardin and Actinonin were purchased by BIOMOL international, SIMP-1 was purchased by Peptide International, Inc.

In vitro assay

The compounds were evaluated for their ability to inhibit the hydrolysis of fluorescence-quenched peptide substrate Mca-Pro-Leu-Gly-Leu-Dpa-Ala-Arg-NH_2 (Biomol, Inc.). The assays were performed in 50 mM HEPES buffer, containing 10 mM $CaCl_2$, 0.05% Brij-35, at pH 7, using 1 nM of MMP-1 catalytic domain and 1 μM of peptide. The enzyme was incubated at 25 °C with increasing concentration of inhibitor and the florescence ($excitation_{max}$ 328 nm; $emission_{max}$ 393 nm) was measured for 3 minutes after the addition of the substrate using a Varian Eclipse fluorimeter. Fitting of rates as a function of inhibitor concentration provided the IC_{50} values. In our experimental conditions with low enzyme concentration and peptide concentration much lower than K_M (the concentration of the substrate that leads to half maximal velocity of the enzymatic hydrolysis reaction), the IC_{50} values provide a good estimate of the dissociation constant of the adduct. The inhibitor N-Isobutyl-N-[4-methoxyphenylsulfonyl]glycyl hydroxamic acid (Biomol, Inc.) was used as control.

NMR measurements

^1H-^{15}N HSQC experiments implemented with the sensitivity enhancement scheme[49] and ^{15}N-NOESY-HSQC spectra[50] were performed on the free MMP-1 catalytic domain and on each protein-ligand adduct. ^{15}N-NOESY-HSQC experiments were acquired with a mixing time of 110 ms and with data sets comprising 256(^1H)×64(^{15}N)×2048(^1H) data points. The NMR spectra were recorded

on Avance 900 Bruker spectrometer, operating at proton nominal frequencies of 900.13 MHz and equipped with a triple resonance cryoprobe. All NMR experiments, recorded at 298 K, were processed using the standard Bruker software (XWINNMR), and analyzed through the XEASY program.[51]

Computer programs

Autodock 3.0.5 was used to predict protein-ligand docking. It uses a Lamarckian genetic algorithm as global optimizer combined with energy minimization as a local search method.[52] Its scoring function is provided by the sum, with empirically determined scaling factors, of a Lennard-Jones 12-6 dispersion/repulsion term, a directional 12-10 hydrogen bond term, a coulombic electrostatic potential, a term related to unfavorable entropy due to restrictions in conformational degree of freedom of the ligand, and a desolvation term. The PDB file was processed by Autodock Tool Kit. Reliable zinc parameters were provided as in Ref. [21]. A box of 70×70×70 points with a grid spacing of 0.375 Å was defined as docking space. The ligands were generated and minimized using semiempirical calculations (AM1 type GAUSSIAN98),[53] and the pdbq files, comprising all protons, were provided to Autodock after all the Gasteiger-Marseli charges[54] were assigned by BABEL. For each run, a maximum number of 28,000 genetic algorithm operations were generated on a single population of 50 individuals. For each ligand a total of 100 docking runs were performed, and the results were ranked according to the docking energy. Crossover, mutation, and elitism weights were set to 0.80, 0.02 and 1, respectively.

All minimization and dynamics calculations were carried out using the program Xplor-NIH.[55,56] The parameter and topology files for the ligands were generated using Xplo2D,[57] the improper angles being manually edited and the dihedral angles being set with force constant equal to zero. Protein electrostatic and van der Waal energy parameters have been evaluated using CHARMM nonbonded parameters.[58]

Acknowledgment

This work was supported by the European Union (contract QLG2-CT-2002-00988 and contract LSHG-CT-2004-512077), by MIUR-FIRB RBNE01TTJW, by MIUR-FISR “Modeling di strutture di metalloproteine e delle interazioni proteina-farmaco e proteina-proteina”, and by Ente Cassa di Risparmio di Firenze.

Reference List

1. Anderson A.C The Process of Structure-Based Drug Design. *Chemistry & Biology* **2003,** *10*, 787-797.
2. Blundell, T. L.; Jhoti, H.; and Abell, C. High-throughput crystallography for lead discovery in drug design. *Nat. Rev. Drug Discov.* **2002,** *1*, 45-54.
3. Lepre, C. A.; Moore, J. M.; and Peng, J. W. Theory and Applications of NMR-Based Screening in Pharmaceutical Research. *Chem. Rev.* **2004,** *104*, 3641-3675.
4. Pellecchia, M.; Becattini, B.; Crowell, K. J.; Fattorusso, R.; Forino, M.; Fragai, M.; Jung, D.; Mustelin, T.; and Tautz L. NMR-based techniques in the hit identification and optimisation processes. *Exp. Opin. Ther. Targets* **2004,** *6*, 597-611.
5. Shuker, S. B.; Hajduk, P. J.; Meadows, R. P.; and Fesik, S. W. Discovering high-affinity ligands for proteins: SAR by NMR. *Science* **1996,** *274*, 1531-1534.
6. Bertini, I.; Calderone, V.; Cosenza, M.; Fragai, M.; Lee, Y.-M.; Luchinat, C.; Mangani, S.; Terni, B.; and Turano, P. Conformational variability of MMPs: beyond a single 3D structure. *Proc. Natl. Acad. Sci. USA* **2005,** *102*, 5334-5339.
7. Stahl, M. and Rarey, M. Detailed Analysis of Scoring Functions for Virtual Screening. *J. Med. Chem.* **2001,** *44*, 1035-1042.
8. Cummings, M. D.; DesJarlais, R. L.; Gibbs, A. C.; Mohan, V.; and Jaeger, E. P. Comparison of Automated Docking Programs as Virtual Screening Tools. *J. Med. Chem.* **2005,** *48*, 962-976.
9. Trosset, J.-Y. and Scheraga, H. A. Reaching the global minimum in docking simulations: A Monte Carlo energy minimization approach using Bezier splines. *Proc. Natl. Acad. Sci. USA* **1998,** *95*, 8011-8015.

10. McCoy, M. A. and Wyss, D. F. Spatial localization of ligand binding sites from electron current density surfaces calculated from NMR chemical shift perturbations. *J. Am. Chem. Soc.* **2002,** *124*, 11758-11763.

11. Lugovskoy, A. A.; Degterev, A. I.; Fahmy, A. F.; Zhou, P.; Gross, J. D.; Yuan, J.; and Wagner, G. A Novel Approach for Characterizing Protein Ligand Complexes: Molecular Basis for Specificity of Small-Molecule BcI-2 Inhibitors. *J. Am. Chem. Soc.* **2002,** *124*, 1234-1240.

12. Pellecchia, M.; Meininger, D.; Dong, Q.; Chang, E.; Jack, R.; and Sem, D. S. NMR-based structural characterization of large protein-ligand interactions. *J. Biomol. NMR* **2002,** *22*, 165-173.

13. Schieborr, U., Vogtherr, M., Elshorst, B., Betz, M., Grimme, S., Pescatore, B., Langer, T., Saxena, K., and Schwalbe, H. How much NMR-Data is required for the Determination of a Protein-Ligand Complex? ChemBioChem . 2005. In Press

14. Woessner, J. F. Jr. Matrix metalloproteinases and their inhibitors in connective tissue remodeling. *FASEB J.* **1991,** *5*, 2145-2154.

15. Yu, A. E.; Hewitt, R. E.; Connor, E. W.; and Stetler-Stevenson, W. G. Matrix metalloproteinases: novel targets for directed cancer therapy. *Drugs Aging* **1997,** *11*, 229-244.

16. Whittaker, M.; Floyd, C. D.; Brown, P.; and Gearing, A. J. Design and Therapeutic Application of Matrix Metalloproteinase Inhibitors. *Chem. Rev.* **1999,** *99*, 2735-2776.

17. Clendeninn, N. J. and Appelt, K. *Metalloproteinase Inhibitors in Cancer Therapy*; Humana Press: Totowa, NJ, 2001.

18. Maeda, A. and Sobel, R. A. Matrix metalloproteinases in the normal human nervous system, microglial nodules, and multiple sclerosis lesions. *J. Neuropathol. Exp. Neurol.* **1996,** *55*, 300-309.

19. Mitton-Fry, R. M.; Anderson, E. M.; Hughes, T. R.; Lundblad, V.; and Wuttke, D. S. Conserved Structure for Single-Stranded Telomeric DNA Recognition. *Science* **2002,** *296*, 145-147.

20. Clore, G. M. and Schwieters, C. D. Docking of Protein-Protein Complexes on the Basis of Highly Ambiguous Intermolecular Distance Restraints Derived from 1HN/15N Chemical Shift Mapping and Backbone 15N-1H Residual Dipolar Couplings Using Conjoined Rigid Body/Torsion Angle Dynamics. *J. Am. Chem. Soc.* **2003,** *125*, 2902-2912.

21. Hu, X. and Shelver, W. H. Docking studies of matrix metalloproteinase inhibitors: zinc parameter optimization to improve the binding free energy prediction. *J. Mol. Graph. Model.* **2003,** *22*, 115-126.

22. MacPherson, L. J.; Bayburt, E. K.; Capparelli, M. P.; Carroll, B. J.; Goldstein, R.; Justice, M. R.; Zhu, L.; Hu, S.; Melton, R. A.; Fryer, L.; Goldberg, R. L.; Doughty, J. R.; Spirito, S.; Blancuzzi, V.; Wilson, D.; O'Byrne, E. M.; Ganu, V.; and Parker, D. T. Discovery of CGS 27023A, a non-peptidic, potent, and orally active stromelysin inhibitor that blocks cartilage degradation in rabbits. *J Med Chem* **1997,** *40*, 2525-2532.

23. Bertini, I.; Calderone, V.; Fragai, M.; Luchinat, C.; Mangani, S.; and Terni, B. X-ray structures of ternary enzyme-product-inhibitor complexes of MMP. *Angew. Chem. Int. Ed.* **2003,** *42*, 2673-2676.

24. Lovejoy, B.; Hassell, A. M.; Luther, M. A.; Weigl, D.; and Jordan, S. R. Crystal Structures of Recombinant 19-kDa Human Fibroblast Collagenase Complexed to Itself. *Biochemistry* **1994,** *33*, 8207-8217.

25. Chen, D. Z.; Patel, D. V.; Hackbarth, C. J.; Wang, W.; Dreyer, G.; Young, D. C.; Margolis, P. S.; Wu, C.; Ni, Z.-J.; Trias, J.; White, R. J.; and Yuan, Z. Actiononin, a Naturally Occurring Antibacterial Agent, Is a Potent Deformylase Inhibitor. *Biochemistry* **2000,** *39*, 1256-1262.

26. BIOMOL Research Laboratories data (www.biomol.com). 2005.

27. Yamamoto, M.; Tsujishita, H.; Hori, N.; Ohishi, Y.; Inoue, S.; Ikeda, S.; and Okada, Y. Inhibition of Membrane-Type 1Matrix Matalloproteinase by Hydroxamate Inhibitors: An Examination of the subsite pocket. *J Med Chem* **1998,** *41*, 1209-1217.

28. van Dijk, A. D. J.; Boelens, R.; and Bonvin, A. M. J. J. Data-driven docking for the study of biomolecular complexes. *FEBS J.* **2005,** *272*, 293-312.

29. Fradera, X. and Mestres, J. Guided Docking Approaches to Structure-Based Design and Screening. *Curr. Top. Med. Chem.* **2004,** *4*, 687-700.

30. Ota, N. and Agard, D. A. Binding Mode Prediction for a Flexible Ligand in a Flexible Pocket using Multi-Conformation Simulated Annealing Pseudo Crystallographic Refinement. *J. Mol. Biol.* **2001,** *314*, 607-617.

31. Taylor, R. D.; Jewsbury, P. J.; and Essex, J. W. A review of protein-small molecule docking methods. *J. Comp. Aided Mol. Design* **2002,** *16*, 151-166.

32. Perola, E.; Walters, W. P.; and Charifson, P. S. A detailed comparison of current docking and scoring methods on systems of pharmaceutical relevance. *Proteins* **2004,** *56*, 235-249.

33. Kellenberger, E.; Rodrigo, J.; Muller, P.; and Rognan, D. Comparative evaluation of eight docking tools for docking and virtual screening accuracy. *Proteins* **2004,** *57*, 225-242.

34. Carlson, H. A. Protein flexibility and drug design: how to hit a moving target. *Curr. Opin. Chem. Biol.* **2002,** *6*, 447-452.

35. Gervasio, F. L.; Laio, A.; and Parrinello, M. Flexible docking in solution using metadynamics. *J. Am. Chem. Soc.* **2005,** *127*, 2600-2607.

36. Maurer, M. C.; Trosset, J. Y.; Lester, C. C.; DiBella, E. E.; and Scheraga, H. A. New general approach for determining the solution structure of a ligand bound weakly to a receptor: structure of a fibrinogen Aα-like peptide bound to thrombin(S195A) obtained using NOE distance constraints and an ECEPP/3 flexible docking program. *Proteins Struct. Funct. Genet.* **1999,** *34*, 29-48.

37. Vogtherr, M. and Fiebig, K. NMR-based screening methods for lead discovery. *EXS* **2003,** *93*, 183-202.

38. Zabell, A. P. Z. and Post, C. B. Docking multiple conformations of a flexible ligand into a protein binding site using NMR restraints. *Proteins Struct. Funct. Genet.* **2002,** *46*, 295-307.

39. Hajduk, P. J.; Mack, J. C.; Olejniczak, E. T.; Park, C.; Dandliker, P. J.; and Beutel, B. A. SOS-NMR: A saturation transfer NMR-based method for determining the structures of protein-ligand complexes. *J. Am. Chem. Soc.* **2004,** *126*, 2390-2398.

40. Hu, X.; Balaz, S.; and Shelver, W. H. A practical approach to docking of zinc metalloproteinase inhibitors. *J. Mol. Graph. Model.* **2004,** *22*, 293-307.

41. Moy, F. J.; Pisano, M. R.; Chanda, P. K.; Urbano, C.; Killar, L. M.; Sung, M. L.; and Powers, R. Asseignments, secondary structure and dynamics of the inhibitor-free catalytic fragment of human fibroblast collagenase. *J Biomol NMR* **1997,** *10*, 9-19.

42. Moy, F. J.; Chanda, P. K.; Chen, J. M.; Cosmi, S.; Edris, W.; Skotnicki, J. S.; Wilhelm, J.; and Powers, R. NMR Solution Structure of the Catalytic Fragment of Human Fibroblast Collagenase Complexed with a Sulfonamide Derivative of a Hydroxamic Acid Compound. *Biochemistry* **1999,** *38*, 7085-7096.

43. Wasserman, Z. R. Making a New Turn in Matrix Metalloprotease Inhibition. *Chemistry & Biology* **2005,** *12*, 143-148.

44. Clore, G. M. Accurate and rapid docking of protein-protein complexes on the basis of intermolecular nuclear overhauser enhancement data and dipolar couplings by rigid body minimization. *Proc. Natl. Acad. Sci. USA* **2000,** *97*, 9021-9025.

45. Clore, G. M. and Bewley, C. A. Using Conjoined Rigid Body/Torsion Angle Simulated Annealing to Determine the Relative Orientation of Covalently Linked Protein Domains from Dipolar Couplings. *J. Magn. Reson.* **2002,** *154*, 329-335.

46. Jain, N. U.; Wyckoff, T. J. O.; Raetz, C. R. H.; and Prestegard, J. H. Rapid Analysis of Large Protein-Protein Complexes Using NMR-derived Orientational Constraints: The 95 kDa Complex of LpxA with Acyl Carrier Protein. *J. Mol. Biol.* **2004,** *343*, 1379-1389.

47. Crowley, P. B.; Otting, G.; Schlarb-Ridley, B.; Canters, G. W.; and Ubbink, M. Hydrophobic interactions in a cyanobacterial plastocyanin-cytochrome *f* complex. *J. Am. Chem. Soc.* **2001,** *123*, 10444-10453.

48. Banci, L.; Bertini, I.; Cavallaro, G.; Giachetti, A.; Luchinat, C.; and Parigi, G. Paramagnetism-based restraints for Xplor-NIH. *J. Biomol. NMR* **2004,** *28*, 249-261.

49. Palmer, A. G., III; Cavanagh, J.; Wright, P. E.; and Rance, M. *J. Magn. Reson.* **1991,** *93*, 151-170.

50. Schleucher, J.; Schwendinger, M.; Sattler, M.; Schmidt, P.; Schedletzky, O.; Glaser, S. J.; Sørensen, O. W.; and Griesinger, C. A general enhancement scheme in heteronuclear multidimensional NMR employing pulsed field gradients. *J. Biomol. NMR* **1994,** *4*, 301-306.

51. Bartels, C.; Xia, T. H.; Billeter, M.; Güntert, P.; and Wüthrich, K. The program XEASY for computer-supported NMR. *J. Biomol. NMR* **1995,** *5*, 1-10.

52. Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; and Olson, A. J. Automated docking using a lamarckian genetic algorithm and empirical binding free energy function. *J. Comput. Chem.* **1998,** *19*, 1639-1662.

53. Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; and Stewart, J. J. P. AM1: a new general purpose quantum mechanical molecular model. *J. Am. Chem. Soc.* **1985,** *107*, 3902-3909.

54. Gasteiger, J. and Marsili, M. Iterative Partial Equalization of Orbital Electronegativity-A Rapid Access to Atomic Charges. *Tetrahedron* **1980,** *36*, 3219-3228.

55. Schwieters, C. D.; Kuszewski, J.; Tjandra, N.; and Clore, G. M. The Xplor-NIH NMR molecular structure determination package. *J. Magn. Reson.* **2003,** *160*, 65-73.

56. Schwieters, C. D. and Clore, G. M. Internal Coordinates for Molecular Dynamics and Minimization in Structure Determination and Refinement. *J. Magn. Reson.* **2001,** *152*, 288-302.

57. Kleywegt, G. J. and Jones, T. A. *Methods Enzymol.* **1997,** *277*, 208-230.

58. MacKerell, A. D. Jr.; Wiòrkiewicz-Kuczera, J.; and Karplus, M. An All-Atom Empirical Energy Function for the simulation of Nucleic Acids. *J. Am. Chem. Soc.* **1995,** *117*, 11946-11975.

Table 1. MMPs residues subjected to significant chemical shifts perturbations are indicated in bold. Chemical shifts for residues in italics are not available.

Seq.	163								171		174												186		204					209
MMP12+NNGH [a]	F	A	R	**G**	**A**	*H*	*G*	D	**D**	..	F	**D**	G	*K*	**G**	**G**	**I**	**L**	**A**	H	**A**	F	G	..	T	T	**H**	**S**	**G**	G
MMP1+NNGH	F	V	R	G	D	H	R	D	N	..	**F**	**D**	G	*P*	G	G	**N**	L	A	H	**A**	F	Q	..	T	N	N	F	R	E
MMP1+Act	**F**	V	R	G	D	H	R	D	N	..	**F**	**D**	G	*P*	G	G	**N**	L	**A**	H	A	F	Q	..	T	N	N	F	R	E
MMP1+SIMP1	**F**	V	*R*	**G**	**D**	H	R	D	**N**	..	**F**	**D**	G	*P*	G	**G**	**N**	*L*	*A*	*H*	**A**	*F*	*Q*	..	T	*N*	**N**	**F**	R	**E**
MMP1+Gal	F	V	R	G	D	H	R	D	**N**	..	F	**D**	G	*P*	**G**	G	*N*	*L*	*A*	**H**	A	F	Q	..	T	N	N	F	R	E

Seq.	210																			230		235								243	249
MMP12+NNGH [a]	**T**	**N**	L	F	L	**T**	**A**	V	H	E	I	G	H	S	L	G	L	G	H	**S**	S	..	V	M	**F**	*P*	**T**	**Y**	K	**Y**	V	**S**
MMP1+NNGH	Y	N	L	H	R	**V**	A	**A**	H	**E**	L	G	H	S	L	G	**L**	**S**	**H**	**S**	T	..	L	M	Y	*P*	**S**	**Y**	T	F	**S**	A
MMP1+Act	Y	N	L	H	R	**V**	A	**A**	**H**	E	**L**	G	H	**S**	L	G	L	**S**	H	S	T	..	**L**	**M**	Y	*P*	S	Y	T	F	S	**A**
MMP1+SIMP1	Y	N	L	H	R	**V**	A	**A**	**H**	E	**L**	G	**H**	**S**	L	G	L	**S**	**H**	S	T	..	**L**	**M**	**Y**	*P*	**S**	**Y**	*T*	*F*	S	A
MMP1+Gal	Y	N	L	H	R	**V**	**A**	A	**H**	E	**L**	G	H	**S**	L	G	L	**S**	**H**	S	T	..	*L*	**M**	**Y**	*P*	S	**Y**	T	*F*	**S**	A

[a] Residue numbers refer to the MMP-1 sequence.

Table 2. Observed NOEs between MMP-1 amide protons and ligand protons

	181	**182**	**184**	**214**	**216**	**221**	**240**	**241**	**242**
NNGH	H1			(H10, H11)	(H10, H11)	H12 H13			
Actinonin	H’	(H3, H4) H6	H1				H’’’ H’ H6	H’’’ (H23, H24)	
Galardin							H’ H’’ (H6, H7)		H’ H’’
SIMP-1	(H17, H18, H19, H20, H21)						H’ H’’	H’’	

Table 3. Predicted H-bonds between protein and ligand nuclei

MMP-12-NNGH	181 LEU HN	NNGH O4 / O3
	182 ALA HN	NNGH O3
	219 GLU HO1	NNGH O1
	NNGH H1	182 ALA O
MMP-1-NNGH	182 ALA HN	NNGH O4
	219 GLU HO1	NNGH O1
	NNGH H1	182 ALA O
MMP-1-Actinonin	181 LEU HN	ACT O3
	219 GLU HO1	ACT O1
	240 TYR HN	ACT O4
	ACT H1	182 ALA O
	ACT N2	238 PRO O
MMP-1-Galardin	181 LEU HN	GAL O3
	182 ALA HN	GAL O3
	219 GLU HO1	GAL O1
	240 TYR HN	GAL O4
	GAL H5	238 PRO O
	GAL H1	182 ALA O
MMP-1-SIMP-1	181 LEU HN	SIMP1 O1
	182 ALA HN	SIMP1 O1
	240 TYR HN	SIMP1 O2
	SIMP1 H8	238 PRO O
	SIMP1 H1	219 GLU OE1
	SIMP1 H23	179 GLY O
	SIMP1 H26	210 TYR OH

Caption to the figures

Figure 1. X-ray structure of the MMP-12-NNGH adduct (PDB: 1RMZ) (A); Residues of MMP-12 affected by chemical shift perturbation upon complexation with NNGH (B); residues of MMP-1 affected by chemical shift perturbation upon complexation with NNGH (C), Actinonin (D), Galardin (E) and SIMP-1 (F) (see Table 2).

Figure 2. Representative structures of the MMP-12-NNGH adduct for the four lowest energy clusters obtained from Autodock (first row), Xplor-NIH (second row), a second Autodock run (third row) and further Xplor-NIH calculations (fourth row). The final validated structures are highlighted.

Figure 3. X-ray structure of the MMP-12-NNGH adduct (A); structures calculated with the proposed protocol of the MMP-12-NNGH adduct (B); structures calculated with the proposed protocol for the adduct of MMP-1 with NNGH (C), Actinonin (D), Galardin (E), SIMP-1 (F). Labels in panels B-F indicate the residue numbers of amino acids exhibiting NOE contacts to the ligands.

Figure 4. Protein-ligand cross peaks observed in the ^{15}N NOESY-HSQC spectrum of the MMP-1-NNGH sample.

Figure 5. Representative structures of the MMP-1-NNGH adduct for the four lowest energy clusters obtained from Autodock (first row), and from Xplor-NIH (second row). At this point convergence was obtained. The final validated structures are highlighted.

Figure 6. Representative structures of the MMP-1-Actinonin adduct for the four lowest energy clusters obtained from Autodock (first row), Xplor-NIH calculations (second row), a second Autodock run (third row) and further Xplor-NIH calculations (fourth row). The final validated structures are highlighted.

Figure 7. Representative structures of the MMP-1-Galardin adduct for the four lowest energy clusters obtained from Autodock (first row), Xplor-NIH calculations (second row), a second Autodock run (third row) and further Xplor-NIH calculations (fourth row). The final validated

structures are highlighted. The configuration of the indole ring is not defined because of lack of experimental restraints and strong energetic preference in Xplor-NIH calculations.

Figure 8. Representative structures of the MMP-1-SIMP-1 adduct for the four lowest energy clusters obtained from Autodock (first row), Xplor-NIH calculations (second row), a second Autodock run (third row) and further Xplor-NIH calculations (fourth row). The final validated structures are highlighted.

Figure 1

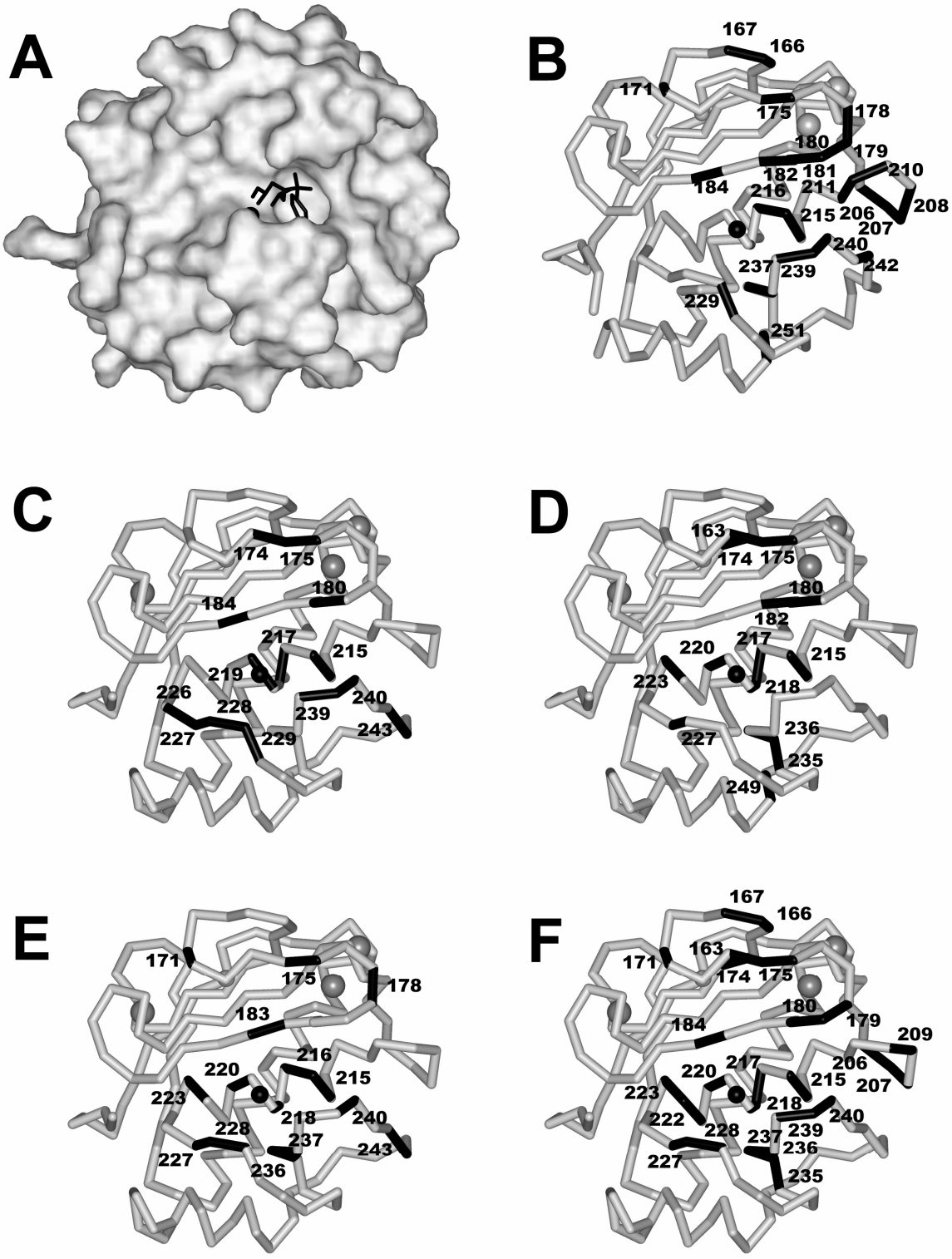

A
B
167
166
171
175
178
180
179
182
181
210
184
216
211
208
215
206
207
240
237
239
242
229
251
C
174
175
180
184
217
215
219
226
228
239
240
227
229
243
D
163
174
175
180
182
220
217
215
223
218
227
236
235
249
E
171
175
178
183
216
220
215
223
218
240
228
237
243
227
236
F
167
166
163
171
174
175
180
179
209
184
220
217
206
215
207
223
218
222
228
237
239
240
236
227
235

Figure 2

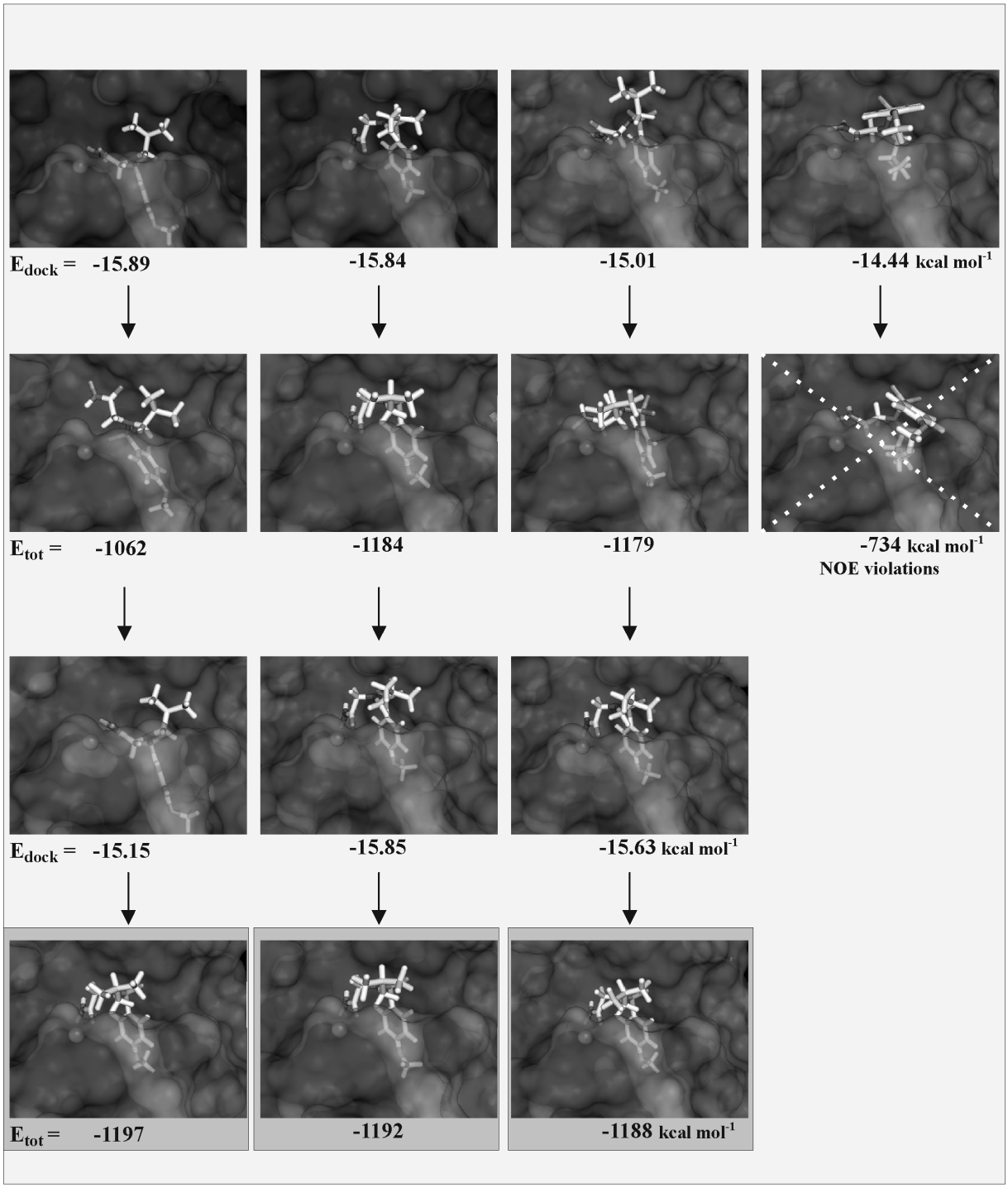

E_{dock} = -15.89
-15.84
-15.01
-14.44 kcal mol^{-1}
E_{tot} = -1062
-1184
-1179
-734 kcal mol^{-1}
NOE violations
E_{dock} = -15.15
-15.85
-15.63 kcal mol^{-1}
E_{tot} = -1197
-1192
-1188 kcal mol^{-1}

Figure 3

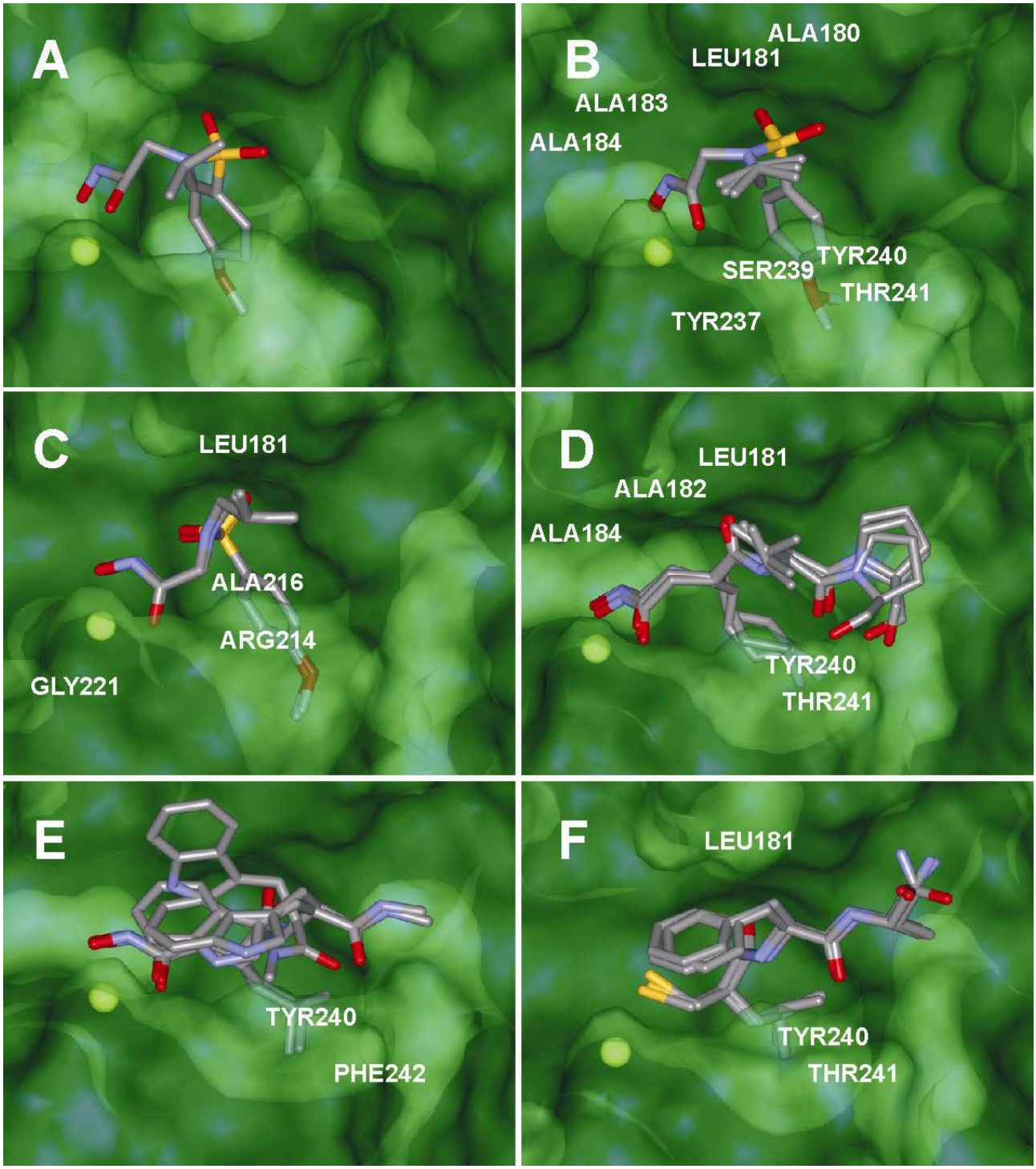

A
B
ALA180
LEU181
ALA183
ALA184
TYR240
SER239
THR241
TYR237
C
LEU181
ALA216
ARG214
GLY221
D
LEU181
ALA182
ALA184
TYR240
THR241
E
TYR240
PHE242
F
LEU181
TYR240
THR241

Figure 4

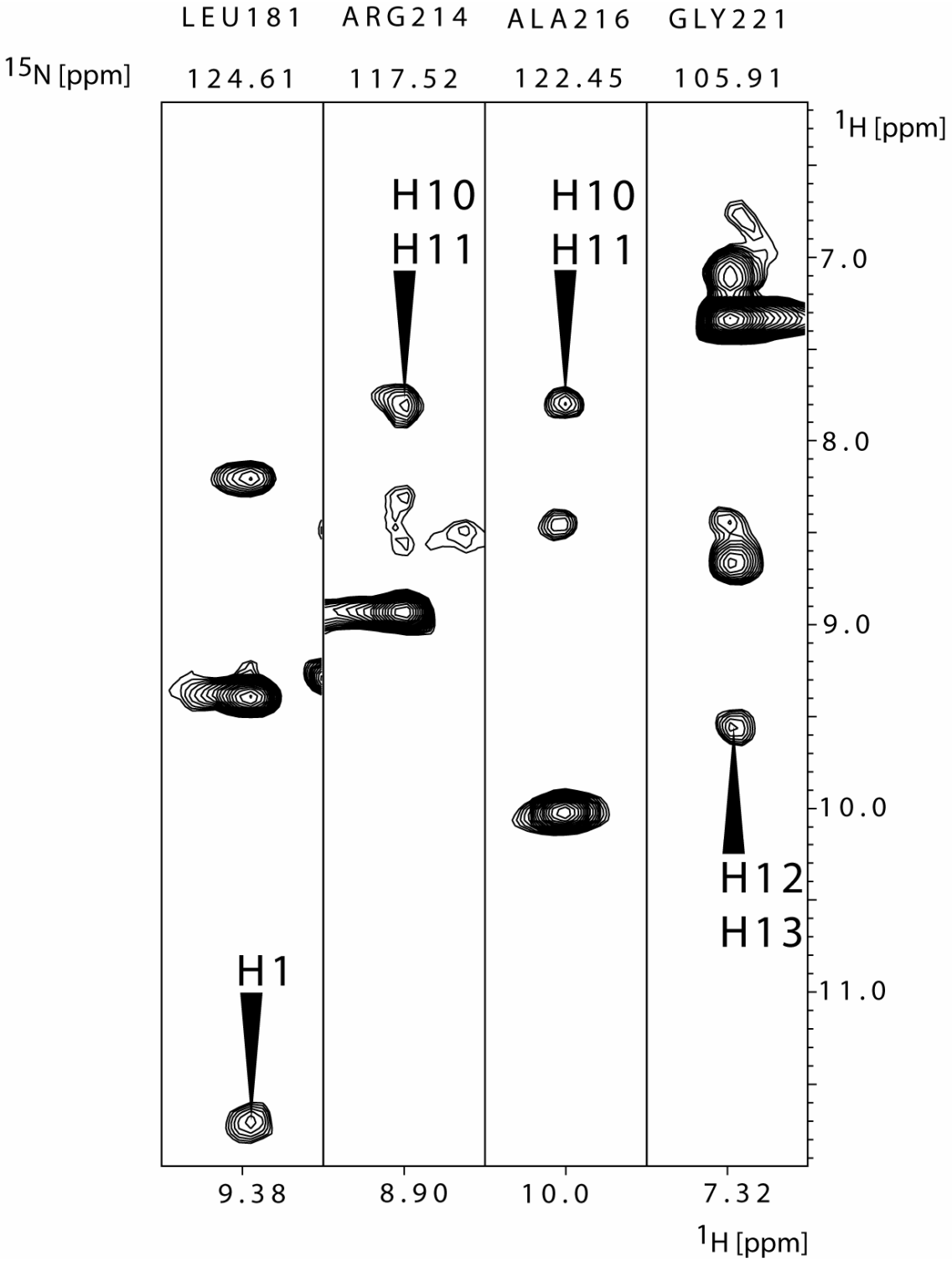

LEU181
ARG214
ALA216
GLY221
15N [ppm]
124.61
117.52
122.45
105.91
1H [ppm]
H10
H11
H10
H11
H1
H12
H13
7.0
8.0
9.0
10.0
11.0
9.38
8.90
10.0
7.32
1H [ppm]

Figure 5

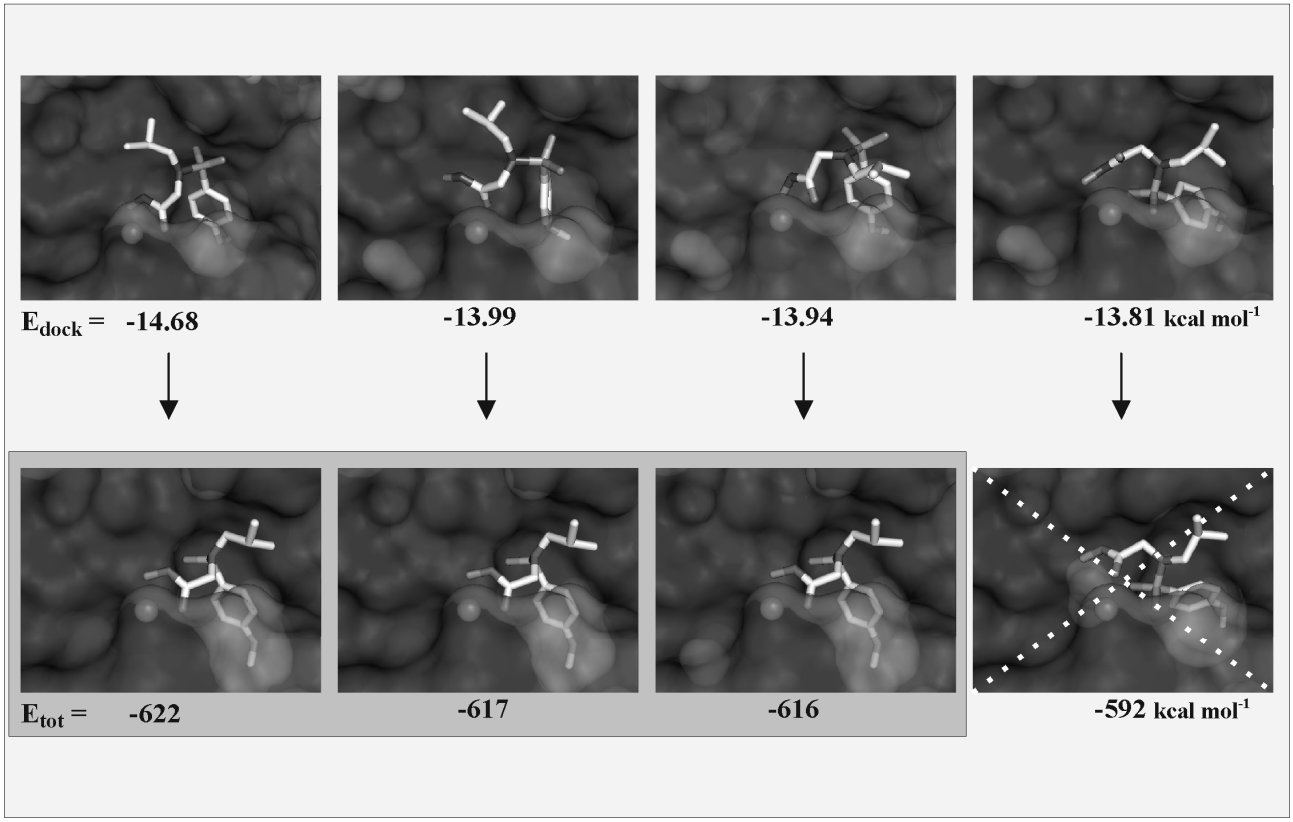

E_{dock} = -14.68
-13.99
-13.94
-13.81 kcal mol^{-1}
E_{tot} = -622
-617
-616
-592 kcal mol^{-1}

Figure 6

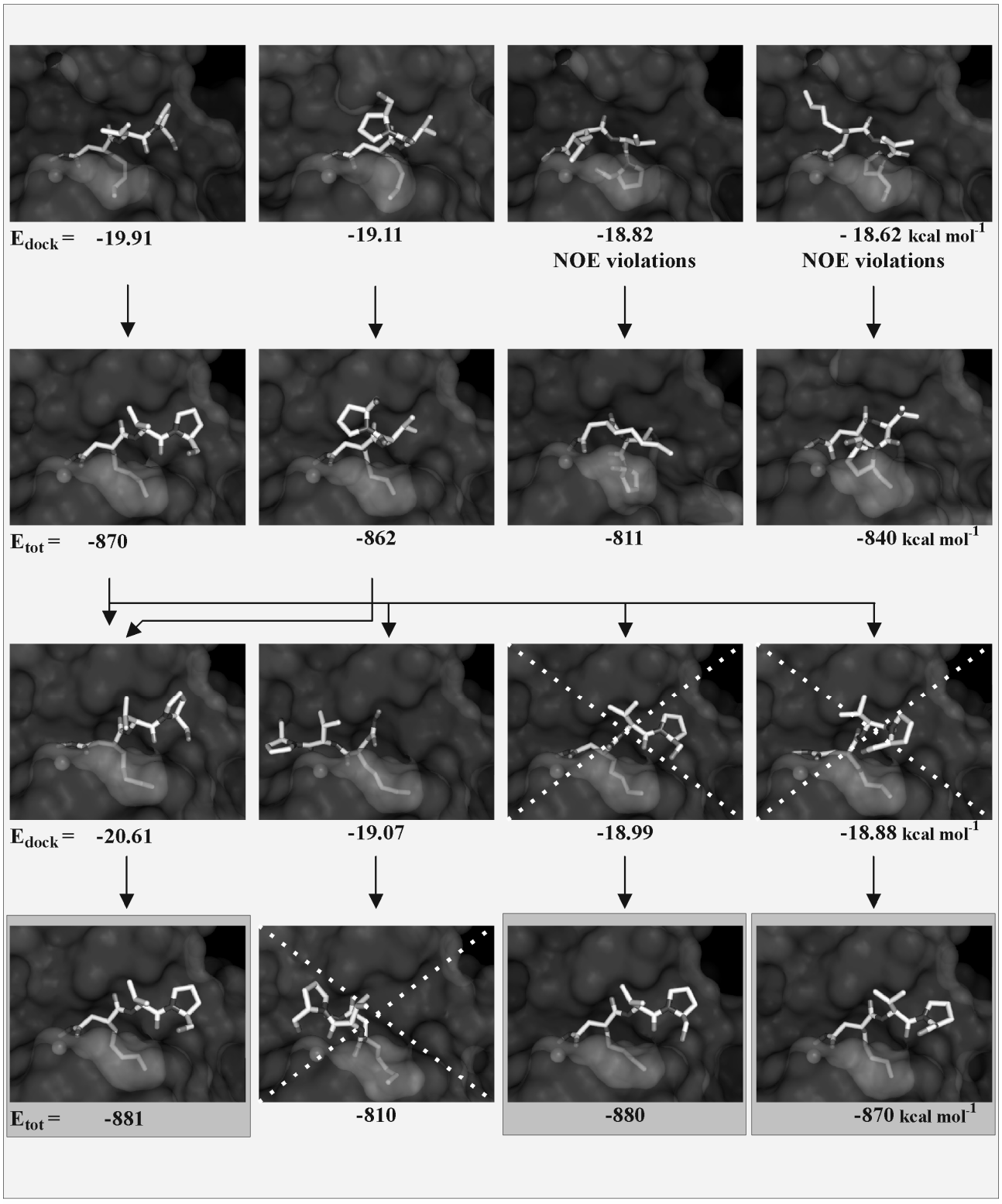

E dock = -19.91
-19.11
-18.82
NOE violations
- 18.62 kcal mol-1
NOE violations
E tot = -870
-862
-811
-840 kcal mol-1
E dock = -20.61
-19.07
-18.99
-18.88 kcal mol-1
E tot = -881
-810
-880
-870 kcal mol-1

Figure 7

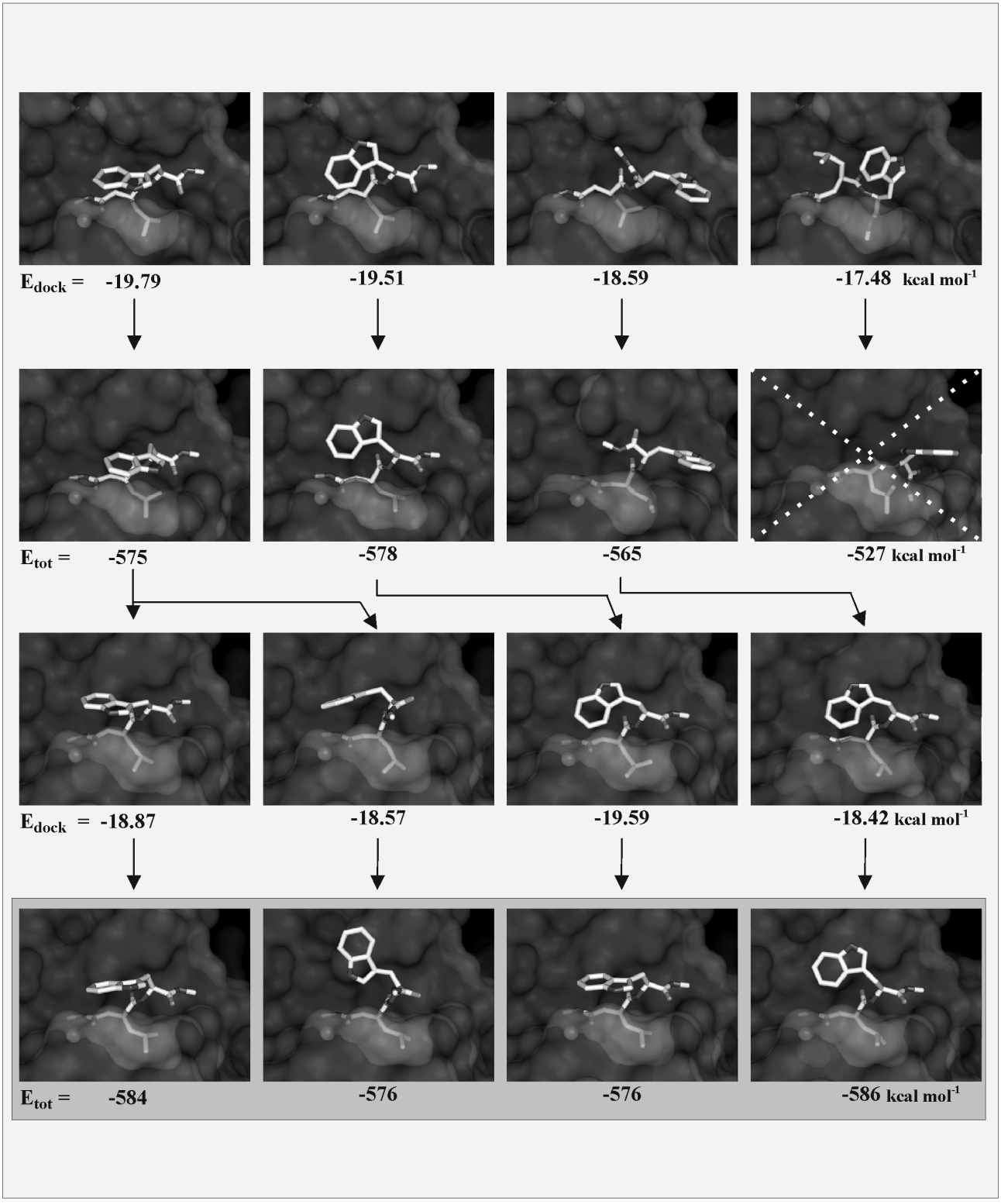
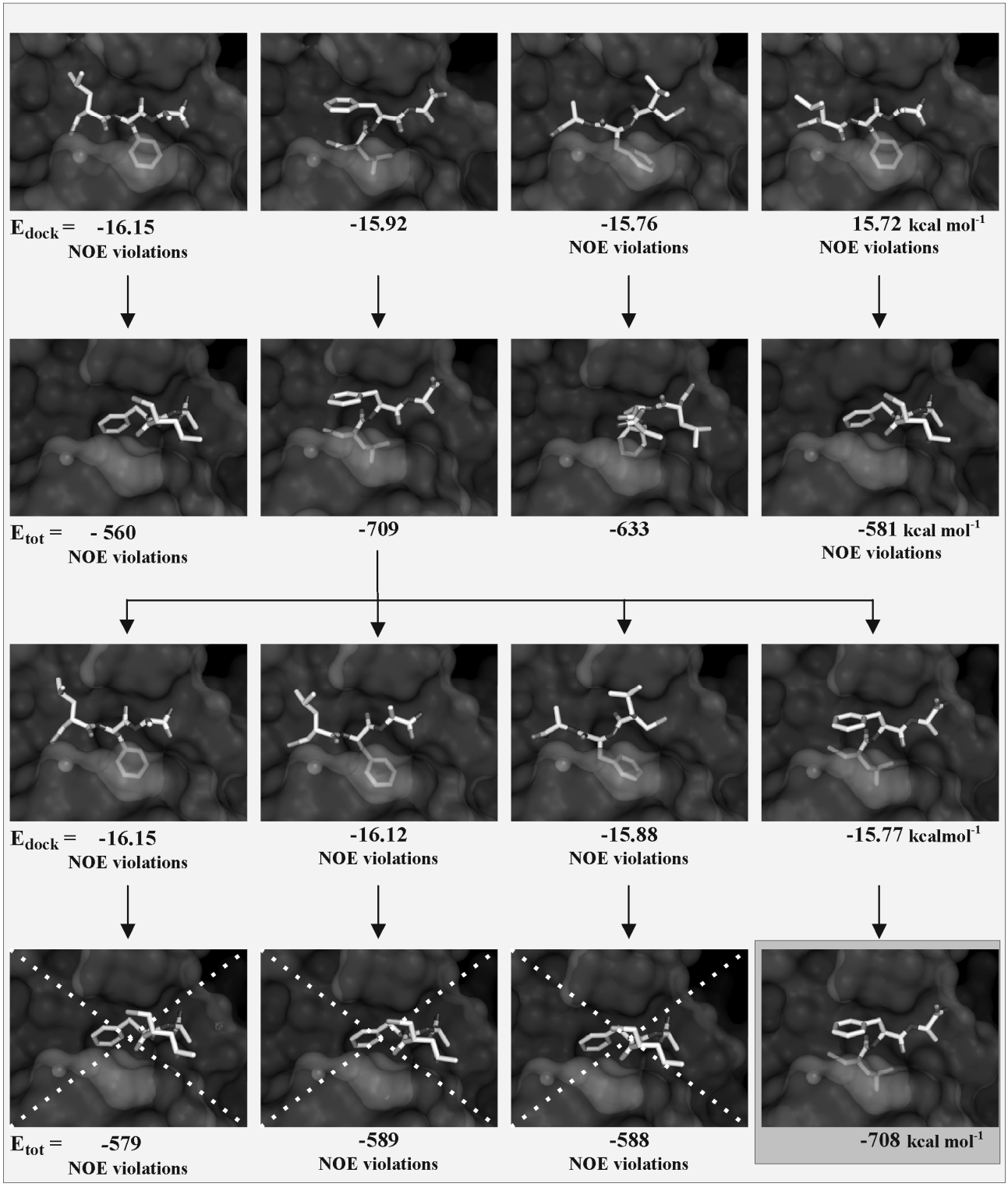

E_{dock} = -19.79
-19.51
-18.59
-17.48 kcal mol^{-1}
E_{tot} = -575
-578
-565
-527 kcal mol^{-1}
E_{dock} = -18.87
-18.57
-19.59
-18.42 kcal mol^{-1}
E_{tot} = -584
-576
-576
-586 kcal mol^{-1}

Figure 8


E dock = -16.15
NOE violations
-15.92
-15.76
NOE violations
15.72 kcal mol-1
NOE violations
E tot = - 560
NOE violations
-709
-633
-581 kcal mol-1
NOE violations
E dock = -16.15
NOE violations
-16.12
NOE violations
-15.88
NOE violations
-15.77 kcalmol-1
E tot = -579
NOE violations
-589
NOE violations
-588
NOE violations
-708 kcal mol-1

Scheme I

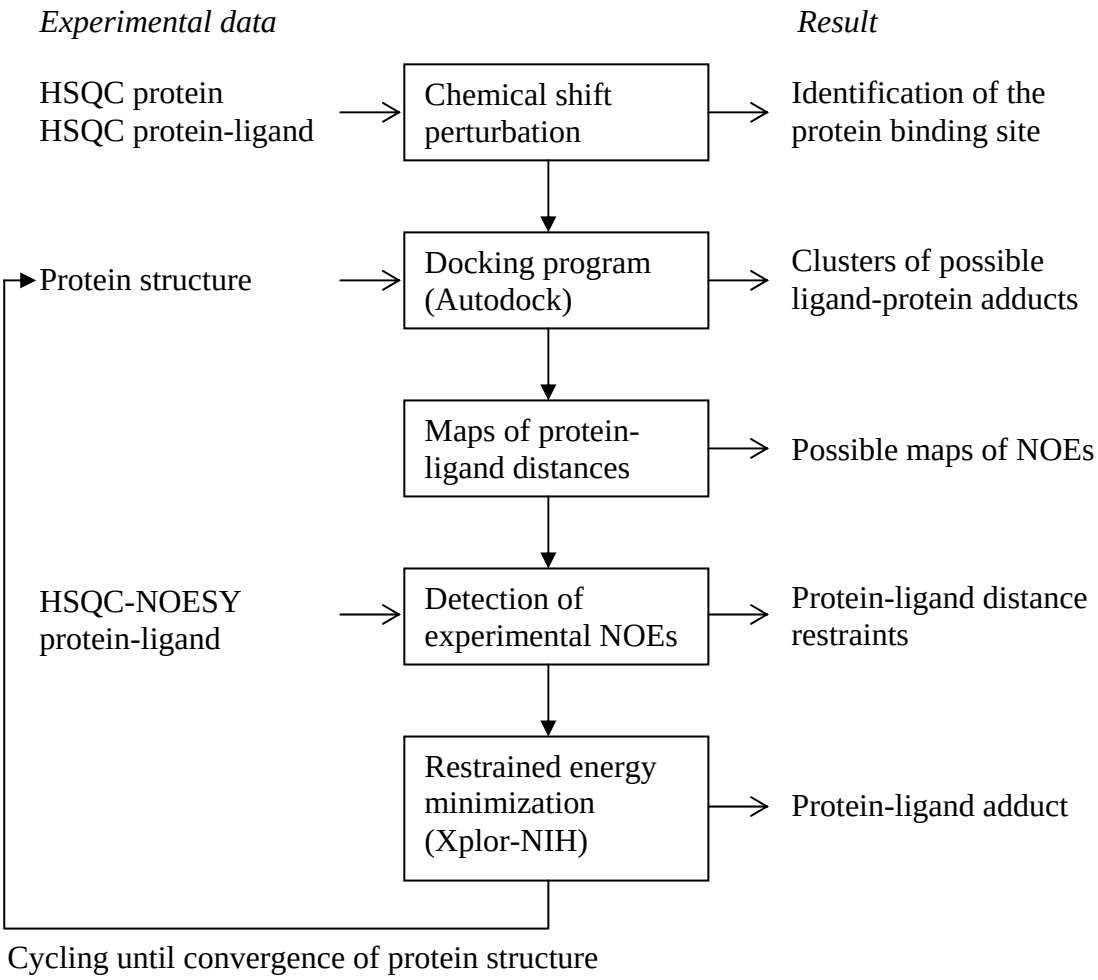
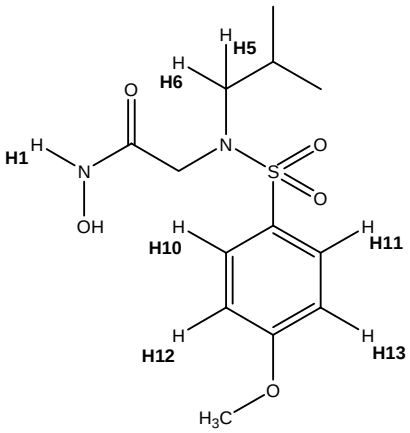
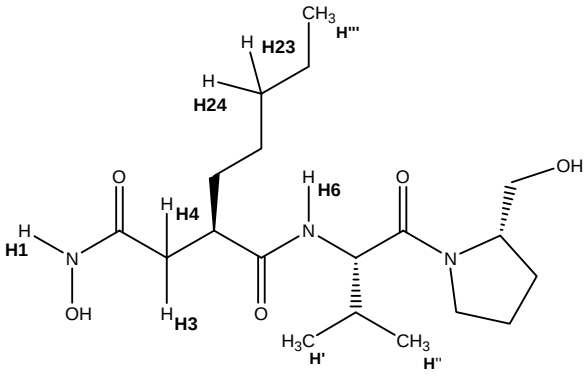

Experimental data
Result
HSQC protein
HSQC protein-ligand
Chemical shift perturbation
Identification of the protein binding site
Protein structure
Docking program (Autodock)
Clusters of possible ligand-protein adducts
Maps of protein-ligand distances
Possible maps of NOEs
HSQC-NOESY protein-ligand
Detection of experimental NOEs
Protein-ligand distance restraints
Restrained energy minimization (Xplor-NIH)
Protein-ligand adduct
Cycling until convergence of protein structure

Chart I.

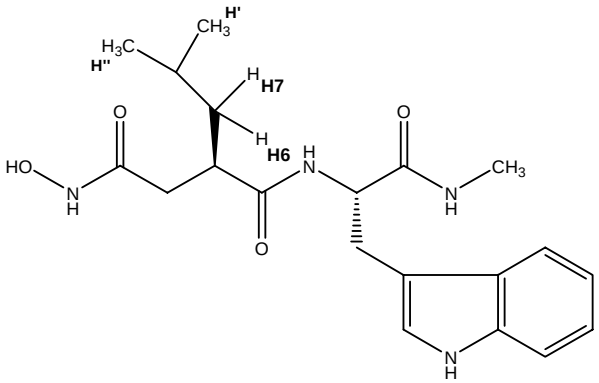
NNGH



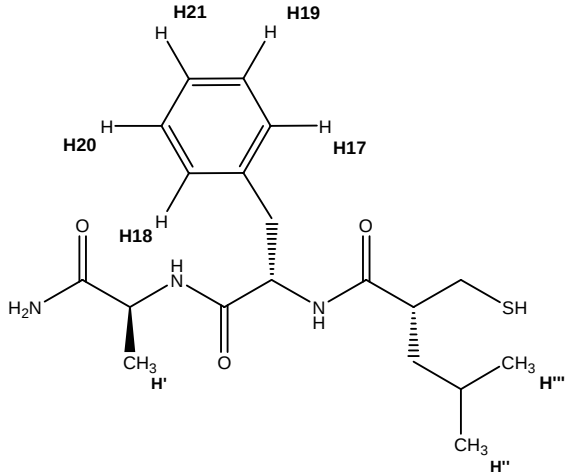
Actinonin



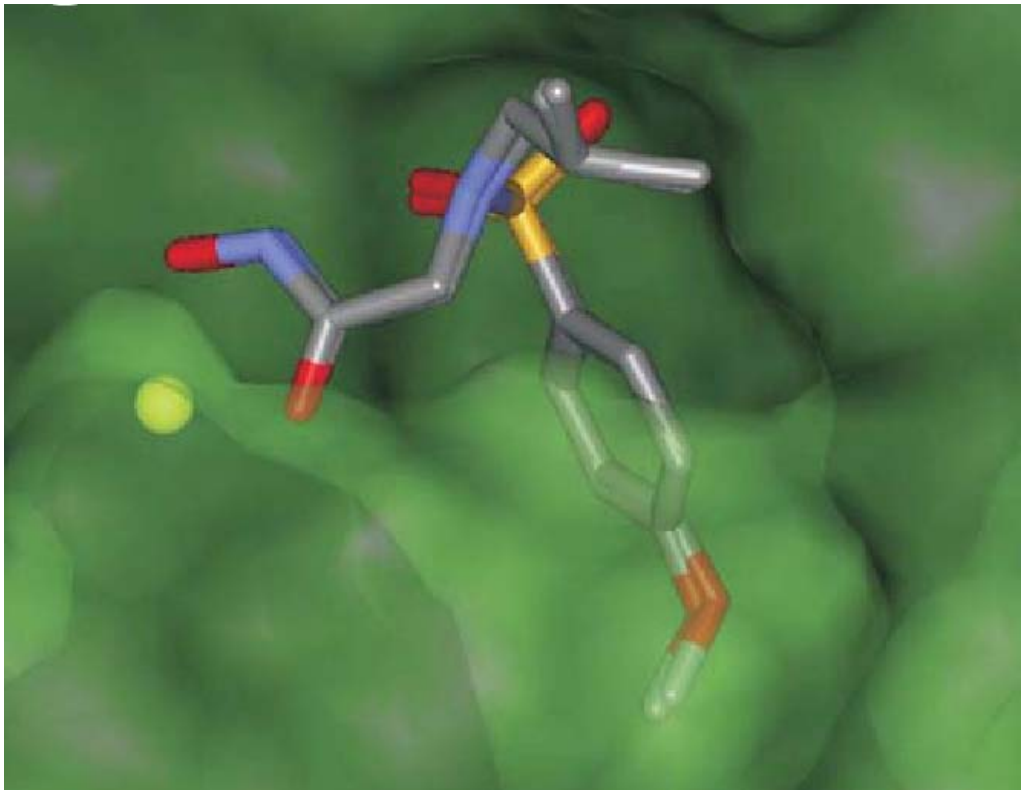
Galardin



SIMP-1



Graphic abstract



4. Conclusions

In the frame of the Course of Doctorate in Structural Biology, my research activity was focused to the development of theoretical techniques that employ molecular dynamics simulations. As a result, several computer protocols have been developed in different fields of research in structural biology. They comprise programs (i) to retrieve the model-free order parameters from molecular dynamics simulations, (ii) to calculate 3-D protein structures through the use of paramagnetism-based restraints, and (iii) to obtain the conformation of protein-ligand adducts using few NMR data. All programs have been tested and applied to different biological systems.

The analysis of 4.5 ns MD simulation of Fe^{2+} and Fe^{3+} redox forms of cyt b_5 in water show that the average orientation of the imidazole ring of histidines coordinated to the iron ion is parallel to the orientation of the B-D meso direction of hem atoms for both orientation state. The dihedral angles of this conformation are consistent with the solution NMR structure. Furthermore one propionate of the heme shows different conformations in the two redox forms, in the oxidized form being bent toward the Fe^{3+} ion because is strongly stabilized by hydrogen bond interaction with the backbone amidic H of Ser-64, in reduced form it is instead screened by solvate sodium ion. Analysis of order parameter show that in some ordered regions, as for instance in the second α-helix (residue 32-43), the N-H bond orientations may have different stable conformations. The transitions between these conformations decrease the value of the order parameter calculated using the whole MD trajectory. To take in account this configuration transiction the system was studied using the block averaging-method(*1*). As a result the agreement with the experimental data is improved (especially in the oxidized form).The program is available for future analysis of different systems.

Paramagnetism-based restraints were included in the Xplor-NIH package for protein solution structure. Xplor-NIH was found to be an efficient tool for exploiting structural information contained in NMR parameters. Paramagnetism-based restraints are paramagnetic relaxation enhancements, pseudocontact shifts, residual dipolar couplings

due to magnetic susceptibility anisotropy, and cross correlation between Curie relaxation and nuclear-nuclear dipolar relaxation. A protocol was developed for their optimal use, as obtained from extensive tests performed on different proteins, and made available through publication on the web. Although such restraints could be already be used in PARAMAGNETIC DYANA(*2*), their introduction in Xplor-NIH was judged to be important to encourage the large number of researchers which use this program to include this type of restraints in their structure calculations.

Finally, I developed a novel protocol for obtaining and validating models of protein-ligand complexes. Two different programs were used in order to take advantage of the respective strengths, and merged into a single protocol. Two proteins belonging to the family of the Matrix Metalloproteinases were used to test and validate the protocol. The quality of the final results depended on the type and number of experimental NOEs but, in all cases, a well defined ligand conformation in the protein binding site was obtained. The strategy proposed ended up begin generally useful for the determination of protein-ligand structure, whenever the structure of the free protein is known and the structural changes upon complexation are minor. This protocol can be useful in a high-throughput approach for structure calculation of protein-ligand adducts.

The physical methods presented in this research aim to elucidate the determination of biomolecular structure, dynamics and binding. Biomolecules interact to form complexes and their dynamics are important for function because they aid interaction and enable access to binding sites. All experimental and computational physical methods presented here measure or provide insights into protein and protein-ligand conformations.

The molecular dynamics simulations have benefited from advances in computer processing power and memory capacities and from novel physical formulations, which permits more realistic simulations. The results available from this work make clear that the computational methods applied to structural biology will play an ever more important role for our understanding of biology in the future.

5. Bibliography

1. P. Koehl, M. Levitt, *Curr. Opin. Struct. Biol.* **9**, 155 (1999).
2. C. L. Brooks, D. A. Case, *Curr. Opin. Struct. Biol.* **13**, 143 (2003).
3. A. T. Brunger, J. Kuriyan, M. Karplus, *Science* **235**, 458 (1987).
4. L. Nillson, G. M. Clore, A. M. Gronenborn, A. T. Brünger, M. Karplus, *J Mol Biol* **188**, 455 (1986).
5. C. L. Brooks, III, M. Karplus, B. M. Petitt, *Proteins: a theoretical perspective of dynamics, structure and thermodynamics* (J. Wiley and Sons, New York, 1988).
6. P. A. Kollman, K. M. Jr. Merz, *Acc. Chem. Res.* **23**, 246 (1990).
7. W. F. Van Gunsteren, *Curr Opin Struct Biol* **3**, 277 (1993).
8. C. M. Dobson, M. Karplus, *Methods Enzymol* **131**, 362 (1986).
9. M. Karplus, *Acc. Chem. Res.* **35**, 321 (2002).
10. T. M. Hansson, C. Oostenbrink, W. F. Van Gunsteren, *Curr Opin Struct Biol* **12**, 190 (2002).
11. W. P. Walters, M. Stahl, M. Murko, *Drug Discov Today* **3**, 160 (1998).
12. T. Langer, R. D. Hoffman, *Curr Pharm Des* **7**, 509 (2001).
13. H. Gohlke, G. Klebe, *Angew. Chem. Int. Ed* **41**, 2644 (2002).
14. H. M. Berman *et al.*, *Nucl. Acids Res.* **28**, 235 (2000).
15. T. L. Blundell, H. Jhoti, C. Abell, *Nat. Rev. Drug Discov.* **1**, 45 (2002).
16. M. Pellecchia *et al.*, *Exp. Opin. Ther. Targets* **6**, 597 (2004).
17. C. A. Lepre, J. M. Moore, J. W. Peng, *Chem. Rev.* **104**, 3641 (2004).
18. A. Giachetti, G. La Penna, A. Perico, L. Banci, *Biophys. J.* **87**, 498 (2004).
19. G. Lipari, A. Szabo, *J. Am. Chem. Soc.* **104**, 4546 (1982).
20. G. Lipari, A. Szabo, *J. Am. Chem. Soc.* **104**, 4559 (1982).

21. L. Banci, I. Bertini, C. Cavazza, I. C. Felli, D. Koulougliotis, *Biochemistry* **37**, 12320 (1998).

22. B. Dangi, J. Blankman, C. J. Miller, B. F. Volkman, R. D. Guiles, *J. Phys. Chem. B* **102**, 8201 (1998).

23. L. Banci *et al.*, *Eur. J. Biochem.* **225**, 715 (1994).

24. I. Bertini, C. Luchinat, *NMR of paramagnetic substances* (Coord.Chem.Rev. 150, Elsevier, Amsterdam, ed. 1, 1996), pp. 1-300.

25. I. Bertini, C. Luchinat, G. Parigi, *Solution NMR of Paramagnetic Molecules* (Elsevier, Amsterdam, 2001).

26. C. D. Schwieters, J. Kuszewski, N. Tjandra, G. M. Clore, *J. Magn. Reson.* **160**, 65 (2003).

27. G. M. Clore, C. A. Bewley, *J. Magn. Reson.* **154**, 329 (2002).

28. L. Banci *et al.*, *J. Biomol. NMR* **28**, 249 (2004).

29. G. M. Morris *et al.*, *J Comp Chem* **19**, 1639 (1998).

30. G. Wagner, *Curr. Opin. Struct. Biol.* **3**, 748 (1993).

31. D. A. Case, *Acc. Chem. Res.* **35**, 325 (2002).

32. G. Wagner, S. Hyberts, J. W. Peng, in *NMR of Proteins*, (Macmillan Press, New York, 1993), pp. 220-257.

33. A. Abragam, *The Principles of Nuclear Magnetism* (Oxford University Press, Oxford, 1961).

34. Indira Chandrasekhar, G.Marius Clore, Attilla Szabo, A. M. Gronenborn, Bernard R.Brooks, *J. Mol. Biol.* **226**, 239 (1992).

35. P. Guntert, *Q. Rev. Biophys.* **31**, 145 (1998).

36. P. Guntert, C. Mumenthaler, K. Wüthrich, *J. Mol. Biol.* **273**, 283 (1997).

37. P. Guntert, *Methods Mol. Biol.* **278**, 353 (2004).

38. A. T. Brunger *et al.*, *Acta Crystallogr D Biol Crystallogr* **54**, 905 (1998).

39. K. Wüthrich, *NMR of Proteins and Nucleic Acids* (Wiley, New York, 1986).

40. H. M. McConnell, R. E. Robertson, *J. Chem. Phys.* **29**, 1361 (1958).

41. R. J. Kurland, B. R. McGarvey, *J. Magn. Reson.* **2**, 286 (1970).

42. I. Bertini, C. Luchinat, G. Parigi, *Progr. NMR Spectrosc.* **40**, 249 (2002).

43. I. Bertini, C. Luchinat, G. Parigi, *Concepts Magn. Reson.* **14**, 259 (2002).

44. V. Gaponenko, A. S. Altieri, J. Li, R. A. Byrd, *J. Biomol. NMR* **24**, 143 (2002).

45. M. Gochin, H. Roder, *Protein Sci.* **4**, 296 (1995).

46. D. L. Turner, L. Brennan, S. G. Chamberlin, R. O. Louro, A. V. Xavier, *Eur Biophys J* **27**, 367 (1998).

47. C. M. Slupsky, R. F. Boyko, V. K. Booth, B. D. Sykes, *J. Biomol. NMR* **27**, 313 (2003).

48. L. Lee, B. D. Sykes, *Biochemistry* **22**, 4366 (1983).

49. L. Banci *et al.*, *J. Biomol. NMR* **12**, 553 (1998).

50. K. Tu, M. Gochin, *J. Am. Chem. Soc.* **121**, 9276 (1999).

51. L. Brennan *et al.*, *J. Mol. Biol.* **298**, 61 (2000).

52. K. L. Bren, J. A. Kellogg, R. Kaur, X. Wen, *Inorg. Chem.* **43**, 7934 (2004).

53. B. J. Goodfellow *et al.*, *Protein Sci* **11**, 2464 (2002).

54. J. R. Tolman, J. M. Flanagan, M. A. Kennedy, J. H. Prestegard, *Proc. Natl. Acad. Sci. USA* **92**, 9279 (1995).

55. L. Banci, I. Bertini, J. G. Huber, C. Luchinat, A. Rosato, *J. Am. Chem. Soc.* **120**, 12903 (1998).

56. J. C. Hus, D. Marion, M. Blackledge, *J Mol Biol* **298**, 927 (2000).

57. P. Tsan *et al.*, *J. Am. Chem. Soc.* **123**, 2231 (2001).

58. I. Bertini, M. Longinetti, C. Luchinat, G. Parigi, L. Sgheri, *J. Biomol. NMR* **22**, 123 (2002).

59. I. Bertini, A. Donaire, C. Luchinat, A. Rosato, *Proteins Struct. Funct. Genet.* **29**, 348 (1997).

60. M. Gueron, *J. Magn. Reson.* **19**, 58 (1975).

61. A. J. Vega, D. Fiat, *Mol. Phys.* **31**, 347 (1976).

62. I. Bertini, C. Luchinat, M. Piccioli, D. Tarchi, *Concepts Magn. Reson.* **6**, 307 (1994).

63. J. Boisbouvier, P. Gans, M. Blackledge, B. Brutscher, D. Marion, *J. Am. Chem. Soc.* **121**, 7700 (1999).

64. J. Bajorath, *Nature Rev. Drug Disco.* **1**, 882 (2002).

65. N. Brooijmans, I. D. Kuntz, *Ann. Rev. Biophys. Biomol. Struct.* **32**, 335 (2003).

66. I. Halperin, B. Ma, H. Wolfson, R. Nussinov, *Protein* **47**, 409 (2002).

67. R. M. Burnett, J. S. Taylor, *Protein* **41**, 173 (2000).

68. R. Norel, S. Lin, H. Wolfson, R. Nussinov, *Biopolymers* **34**, 933 (1994).

69. R. Norel, H. Wolfson, R. Nussinov, *Comp. Chem. High Throughput Screen* **2**, 177 (1999).

70. J. S. Dixon, C. M. Oshiro, *J. Comput. Aided Mol. Des.* **9**, 113 (1995).

71. G. Jones, P. Willet, R. C. Glen, A. R. Leach, R. Taylor, *J Mol Biol* **267**, 727 (1997).

72. J.-Y. Trosset, H. A. Scheraga, *Proc. Natl. Acad. Sci. U. S. A* **95**, 8011 (1995).

73. G. Jones, P. Willet, R. C. Glen, *J Mol Biol* **245**, 53 (1995).

74. T. J. A. Ewing, S. Makino, A. G. Skillman, I. D. Kuntz, *J. Comp. Aided Mol. Design* **15**, 411 (2001).

75. B. Kramer, M. Rarey, T. Lengauer, *Proteins* **37**, 228 (1999).

76. S. J. Weiner, P. A. Kollman, D. T. Nguyen, D. A. Case, *J. Comp. Chem.* **7**, 287 (1986).

77. J. H. Holland, *Ann Arbor, MI: University of Michigan Press* (1975).

78. D. E. Clark, *New York: Wiley-VCH* (2000).