Università degli Studi di Firenze

Facoltà di Scienze Matematiche, Fisiche e Naturali

Dottorato in Scienze Chimiche

Ciclo XXII

Metodologie NMR per la caratterizzazione strutturale e dinamica di biomolecole in soluzione

Mirko Mori

Tutor

Prof. Mario Piccioli

Coordinatore

Prof. Gianni Cardini

CONTENTS

1. Introduzione 5

2. Structure determination through NMR spectroscopy 9

2.1 Precision and accuracy of NMR structures 10

2.2 Progress in restraints 11

2.2.1 Residual dipolar coupling 13

2.3 How to assess the quality of NMR structures 15

3. NMR characterization of the slow dynamics in biomacromolecules 23

3.1 NMR relaxation and molecular dynamics 24

3.2 Relaxation techniques for slow dynamics: Chemical Shift Modulation 26

4. Overview of the investigated proteins 30

4.1 Copper, Zinc Superoxide dismutase 30

4.2 Matrix Metalloproteinases 32

4.3 Calbindin D_{9k} 33

5. The Solution Structure of the Monomeric Copper, Zinc Superoxide Dismutase from Salmonella enterica: Structural Insights To Understand the Evolution toward the Dimeric Structure 37

6. Towards structural dynamics: fluctuations in proteins monitored by chemical shift modulations and direct detection of C'N multiple-quantum relaxation rates 54

Acknowledgements

Il primo sincero ringraziamento di questa tesi va al mio tutore di Dottorato, il Prof. Mario Piccioli, che mi ha accompagnato in tutti questi tre anni rendendo questa esperienza molto interessante e allo stesso tempo piacevole, non solo dal punto di vista scientifico ma, non per ultimo, anche sotto il profilo umano.

Desidero sentitamente ringraziare il Prof. Ivano Bertini per tutti gli insegnamenti ricevuti in questi anni sul significato di "fare Scienza" e per avermi dato l'opportunità di conoscere un mondo senz'altro affascinante. Ringrazio la Prof. Lucia Banci e il Prof. Claudio Luchinat per la loro disponibilità e le preziose indicazioni da loro ricevute.

Un ringraziamento speciale va a tutti gli amici del CERM con i quali abbiamo trascorso dei bei momenti, contribuendo non poco a rendere particolarmente piacevole questa esperienza. Quindi, ringrazio la consolidata comitiva delle "allegre pause pranzo", Stefano "fra le nuvole", Manuele "il piemontese", Sara "la baronessa", il "grande e grosso" Angiolone e, ultima arrivata, Maria Antonietta, "Napoli". Poi, tutti i compagni di ufficio che sono passati nei tre anni che hanno contribuito a far diventare quelle fredde quattro mura di cartongesso nella fantastica "Stanza 4",in particolare Leonardo "la spalla perfetta", Serena "la risata assordante", Hembram "l'indiano enigmatico" e, ultimamente, il simpatico duetto "CristiNepi"….e poi tanti altri, Chiara, Sara, Valentina, Eva, Mirco, Daniela, Camilla……. con cui ho lavorato insieme o semplicemente preso un caffè.

Voglio ringraziare anche Francesca, Simone e Marco per la loro continua e quotidiana disponibilità tutte le volte che ne ho avuto bisogno.

Desidero inoltre ringraziare tutti i tecnici del CERM, Massimo Lucci, Fabio Calogiuri, Enrico Morelli, Leonardo Gonnelli, Marco Allegrozzi e Rebecca Del Conte per la loro disponibilità e il loro aiuto.

Infine, ringrazio tutte le persone con cui ho collaborato in questi tre anni senza le quali questo lavoro non sarebbe stato possibile, in particolare Beatriz Jiménez, Marco Sette e Daniel Abergel.

A Silvia,

una persona speciale che mi sta vicino e mi sopporta da più di 10 anni…..

Capitolo 1

1. *Introduzione*

Questo lavoro di tesi si inserisce nell'ampio settore delle applicazioni della spettroscopia di Risonanza Magnetica Nucleare (RMN) per lo studio di sistemi di interesse biologico. In particolare, l'obiettivo di questa tesi è l'applicazione e lo sviluppo di metodologie innovative volte alla caratterizzazione strutturale e dinamica di macromolecole biologiche in soluzione.

La conoscenza della struttura di molecole biologiche rappresenta un importante informazione per la studio dei vari processi biologici. La RMN e la cristallografia a raggi X sono le due maggiori tecniche utilizzate per la caratterizzazione strutturale a livello atomico di proteine, acidi nucleici e l'identificazione delle interazioni con altre molecole biologiche. L'utilizzo della RMN per la determinazione della struttura tridimensionale di biomolecole in soluzione è ormai noto da circa 25 anni, nel corso dei quali il numero delle strutture risolte via RMN è costantemente aumentato. Su di un totale di circa 56000 strutture depositate nel Protein Data Bank (PDB), più di 7600 sono state determinate via RMN. Oggigiorno, entrambe le tecniche hanno un ruolo chiave nei vari progetti di Proteomica Strutturale e, sempre più spesso, un approccio di tipo complementare è fondamentale nella determinazione strutturale delle molecole biologiche. In questo settore, l'utilizzo della RMN diventa fondamentale, non solo quando il sistema non può essere cristallizzato, ma anche per sfruttare le uniche potenzialità offerte dalla spettroscopia di RMN in soluzione. Alcuni esempi che si aggiungono alla determinazione strutturale sono la possibilità di studiare la dinamica molecolare, gli equilibri conformazionali, i fenomeni di *binding* con altri ligandi, gli aspetti cinetici e termodinamici delle interazioni tra macromolecole, la caratterizzazione di polipeptidi funzionali non strutturati e la loro transizione in complessi attivi. Infine, la RMN offre la possibilità di effettuare esperimenti in soluzione acquosa in condizioni simili a quelle fisiologiche. Nel **Capitolo 2** saranno presentate e discusse le

problematiche relative alla determinazione della struttura tridimensionale di molecole biologiche in soluzione via RMN.

Inizialmente, la conoscenza della struttura di macromolecole biologiche era considerata come l'ultimo stadio dello studio di un sistema, in cui l'analisi biochimica della molecola in esame era basata prevalentemente sullo studio della struttura ottenuta, come se fosse statica, senza considerare nessun fenomeno di dinamica ad essa correlata. Poi, con l'aumentare delle conoscenze chimico-biologiche, è stato evidenziato come la struttura rappresenti, sempre più spesso, un primo dato fondamentale di conoscenza del sistema per ulteriori studi successivi. Infatti, innumerevoli processi biologici sono direttamente correlati a cambi conformazionali di molecole biologiche come nei fenomeni di *folding* e *unfolding*, ricognizione molecolare, catalisi ed interazione con altri ligandi. La dinamica interna del sistema è quindi essenziale per la funzione della molecola stessa. La possibilità di studiare il sistema in soluzione in condizioni simili a quelle fisiologiche è una peculiarità della RMN, ciò permette di analizzare le proprietà dinamiche della molecola in esame a livello atomico che comprende moti su diverse scale del tempo, dai nano ai secondi. Dato che molti processi biologici avvengono nell'intervallo che va dai micro ai millisecondi, si presuppone che lo studio della dinamica interna in questo intervallo di tempo (*slow dynamics*) sia correlato all'attività biologica. La caratterizzazione della *slow dynamic* diventa così di fondamentale importanza per la comprensione della funzione biologica delle macromolecole, proteine e acidi nucleici.

Per raggiungere tali obiettivi, in questo lavoro di tesi è stato utilizzato un nuovo approccio che prevede l'utilizzo di esperimenti di RMN innovativi. In particolare, è stata sfruttata la dipendenza della velocità di rilassamento della coerenza multiplo quanto (MQ) fra il carbonio carbonilico (^{13}C') e l'azoto ammidico (^{15}N) del piano peptidico delle proteine e la presenza di moti di dinamica lenta del *backbone* della molecola. Per misurare la velocità di rilassamento MQ abbiamo implementato una nuova sequenza di impulsi basata sull'osservazione diretta del nucleo a bassa sensibilità ^{13}C. Questa metodologia è resa possibile grazie alla disponibilità presso il laboratorio CERM di strumentazione in alta risoluzione a campo variabile da 200 a 900 MHz, in parte ottimizzata per l'osservazione di tali nuclei.

La sequenza è la versione modificata del già noto esperimento CON, dove si osserva la correlazione tra il carbonio carbonilico del residuo *i* e l'azoto ammidico del residuo *i+1* , scalarmente accoppiato al carbonio. Rispetto all'approccio standard che prevede l'osservazione classica del nucleo ^{1}H in un esperimento tipo HNCO, l'esperimento CON presenta alcuni vantaggi. Questo tipo di esperimento permette l'osservazione diretta dei nuclei interessati dal fenomeno che vogliamo studiare, inoltre possiamo usufruire dei tipici vantaggi dell'osservazione diretta del ^{13}C, quali la maggiore dispersione in chemical shift dei segnali (quindi minore *overlap* degli stessi) e la possibilità di monitorare zone soggette a scambio chimico che preclude l'osservazione dei segnali ^{1}H-^{15}N in un piano HSQC. Nel **Capitolo 3** saranno affrontate le problematiche inerenti lo studio della dinamica interna di molecole biologiche, con particolare attenzione alla caratterizzazione dei moti di *slow dynamics*.

I sistemi che abbiamo studiato in questo lavoro sono l'enzima Cu,Zn-superossido dismutasi (CuZnSOD) da *Salmonella choleraesuis*, il dominio catalitico della Metallo Proteinasi di Matrice 12 (MMP12) in presenza dell'inibitore NNGH (acido *N*-isobutil-*N*-[4-metossifenilsulfonil] glicil idrossammico) e la proteina Calbindina D_{9k} (Calb). Una presentazione dei tre sistemi è riportata nel **Capitolo 4**.

La determinazione della struttura dell'enzima CuZnSOD è stata preliminare allo studio della dinamica interna ed a una analisi spettroscopia e strutturale dei fenomeni di idratazione. L'interesse per questa proteina deriva del fatto che è la prima struttura in soluzione risolta via RMN di un enzima monomero allo stato nativo, appartenente alla classe delle superossido dismutasi. La loro funziona biologica è di primaria importanza per gli organismi viventi in quanto catalizzano la reazione di dismutazione del radicale superossido, tossico per le cellule, in ossigeno e acqua ossigenata. Questi enzimi sono stati scoperti circa 40 anni fa e, ancora oggi, sono una delle classi di enzimi più studiata in termini di proprietà biofisiche, biochimiche e loro coinvolgimento in processi degenerativi cellulari, quali FALS (*Familial Amyotrophic Lateral Sclerosis*). Gli studi effettuati su proteine isolate da organismi eucarioti e procarioti hanno evidenziato l'esistenza di un alto grado di diversità strutturale tra i due domini. Gli enzimi isolati da eucarioti sono tutti omodimeri, mentre fra i procarioti sono presenti anche esempi di monomeri. Inoltre, mentre nel primo dominio è conservata un'alta omologia strutturale fra le varie proteine, nei procarioti si osserva una maggiore variabilità strutturale sia in termini di organizzazione del sito attivo che diversa associazione delle sub-unità, nel

caso di dimeri. Le differenze e le similitudini tra questo enzima e gli altri isoenzimi batterici, così come l'isoenzima umano omodimero, assai ben conosciuto e studiato, sono state studiate e discusse nel **Capitolo 5**.

Nel **Capitolo 6** sono presentati gli studi di dinamica che cadono nell'intervallo dai micro ai millisecondi effettuati sulle proteine Calb, CuZnSOD e MMP12. La scelta delle tre proteine è dovuta alla volontà di studiare il comportamento di diversi tipi di *folding* dal punto di vista della dinamica interna: un sistema ad α-elica di tipo EF-hand, un barile di tipo β e un sistema misto α+β (2α-eliche + 5β-foglietti), rispettivamente.

Grazie all'innovativo approccio decritto in precedenza (vedi Capitolo 3), abbiamo misurato le velocità di rilassamento delle coerenze MQ per il $^{13}C'$ e ^{15}N del piano peptidico delle proteine. Le fluttuazioni della componente isotropa del chemical shift (CSM, *chemical shift modulation*) dei due nuclei sono correlate ai movimenti del *backbone* della proteina nell'intervallo dei micro- millisecondi e contribuiscono alla diversa larghezza di riga delle coerenze zero quanto (ZQ) e doppio quanto (DQ). Dalla misura delle velocità di rilassamento MQ è possibile fattorizzare il solo contributo causato dalla velocità di cross-rilassamento dovute alle modulazione del *chemical shift* ($R^{CSM/CSM}$) dei due nuclei interessati al fenomeno. L'analisi dei dati di rilassamento evidenzia la presenza di regioni affette da un diverso grado di *slow dynamic* in tutti e tre i sistemi studiati. In particolare, il risultato più interessante è stato la correlazione tra le velocità di cross-rilassamento $R^{CSM/CSM}$ e la struttura secondaria della proteina. Regioni strutturate a α-elica sono affette da fluttuazione anti-correlanti del *chemical shift* ($R^{CSM/CSM} <0$), mentre regioni strutturate a β-foglietto sono interessate da fluttuazioni correlanti del *chemical shift* ($R^{CSM/CSM} >0$). Inoltre, l'andamento dei valori di rilassamento in funzione dell'angolo diedro del *backbone* ψ mostra una distribuzione dei dati tipo curva di Karplus. Sebbene i dati siano affetti da un largo *scattering*, che preclude una possibile parametrizzazione degli stessi, questo trend è osservato chiaramente in modo non ambiguo per tutte e tre i sistemi studiati. La diretta correlazione tra una velocità di rilassamento ed un parametro strutturale che definisce la geometria locale della proteina, come l'angolo diedro ψ, rappresenta un risultato innovativo, dal momento che in letteratura non esistono altri esempi di questa tipologia.

Chapter 2

2. Structure determination through NMR spectroscopy

The NMR restraints used in structure determination are loose in nature and too few with respect to the degrees of freedom of a protein, to produce a well resolved structure. Typically, in a protein structure the restraints are of the order of 10-20 independent interatomic distances per aminoacid (obtained from NOE measurements) plus some dihedral angles and some "average" atom-atom vector directions. These restraints produce a structure whose resolution corresponds at most to an X-ray resolution of 2 Å. Furthermore, proteins have intrinsic mobility and movements occurring below milliseconds provide average restraints. The sensitivity of NMR to motions is a drawback for the obtainment of high resolution structures, but it is also a source of information on the internal dynamics of proteins and nucleic acids. NMR experiments and analysis approaches have been developed to obtain these information.

One possible drawback in protein structural determination by NMR is that the assignment of the NMR signals is still mainly operator-dependent and errors may be possible. If a signal is mis-assigned, wrong distance restraints are obtained. Sometimes no structure could be determinate with a wrong restraint and therefore the error becomes apparent, but it is also possible that the error leads to wrong local conformation. The response of the scientific community to this challenge is on the one hand, the development of software for automated assignment, but at the moment the errors are many, and on the other the development of tools to assess the reliability of the resulting structure on the basis of statistical procedures [1].

For NMR structures, it is customary to present NMR structures as an ensemble of conformers, called family, each component of which is consistent with the experimental data within a given tolerance. The agreement between experimental restraints and those measured on the calculated conformer is estimated and the conformers within a selected threshold for this agreement are selected (target function). The scattering among the conformers of the family is evaluated by the root mean square deviation (RMSD) from

the average conformer. A backbone RMSD value of 0.5 Å, which corresponds to an excellent NMR structure, corresponds to an X-ray resolution of about 2 Å.

Another drawback in NMR structure determination is the time it takes to solve a structure. The software for structure determination is continuously improving and the time needed has decreased and will further decrease. Attempts to standardize the programs and to allow NMR user friendly approaches to even non expert users are in progress (http://www.enmr.eu). However, the NMR measurement times are still very long, i.e. of the order of several weeks. Therefore attempts are also being made to reduce the experimental time by new pulse sequences (fast NMR) [2,3] or by innovative NMR data acquisition schemes (non-linear time sampling [4], projection NMR [5], Hadamard-type [6], spatial frequency encoding, extensive aliasing [7] and multi-way decomposition of NMR spectra [8,9]. Moreover, a structure calculation program, Chemical-Shift-ROSETTA (CS-ROSETTA), is available based on empirical correlations gained from the mining of protein chemical shifts deposited in the BMRB [10], in conjunction with the known corresponding 3D structures. CS-ROSETTA [11] is a robust protocol to exploit this relation for de novo protein structure generation, using as input parameters the $^{13}C^{\alpha}$, $^{13}C^{\beta}$, $^{13}C'$, ^{15}N, $^{1}H^{\alpha}$ and $^{1}H^{N}$ NMR chemical shifts. The generated models are rescored based on the difference between the back-calculated chemical shifts of the generated models and the input chemical shifts.

2.1 Precision and accuracy of NMR structures

Any experiment, even if carefully performed, has errors associated with it. These errors are of two distinct nature: systematic and random. Systematic errors relate to the accuracy of the model. They may occur at the stage of spectral interpretation where the individual NMR signals are assigned to active NMR atoms. Random errors depend on how precisely a given measurement can be made. All measurements contain errors at some degree of precision. If a model is essentially correct, the size of random errors will determine how precise the model is. The distinction between accuracy and precision is a particularly important one in NMR.

The accuracy is a measure of how close is the determined structure to the true structure and can only be obtained when a ‘gold standard’ is available, like in the case of a

simulated data set. For real NMR structures the accuracy is often calculated with respect to a reference X-ray structure of the same molecule [12-14]. An indicator of the accuracy of NMR structures is provided by their validation against independently known structural restraints such as dipolar couplings [15] which were not directly used in the calculations. The agreement between calculated and measured dipolar couplings is often expressed in terms of a Q-factor. The better the agreement, the lower the Q-factor. A suitable way to evaluate solution structures is the calculation of the NMR R-factor, that directly compares an experimental NMR NOESY spectrum with the corresponding spectrum back-calculated for each conformer of the family or for a subset of them [16-18].

The precision, expressed as the RMS of the deviations in position of an atom in one conformer and that in another in the NMR family, is commonly used as an indication of how well the structure has been refined. Normally, families of conformers are generated by selecting the lowest target function conformers from the ensemble calculated with the experimental data. The RMSD within the family therefore depends not only on the quality and quantity of available data but also on the procedure used to select the conformers from the calculated ensemble. In literature several works on the relationship between precision and accuracy of the ensemble of calculated conformers are available and it has been often found that the search for a too high precision reduces the accuracy [19,20]. In order to properly analyze the available structural restraints, they can be slightly relaxed. Vriend et al described an approach (re-sampling method) to improve the sampling and representation of the conformational space defined by the experimental restraints. Their approach showed that, for all test cases, the calibration of NMR restraints used in the structure calculation can be loosen still obtaining a good fit of the experimental data, maintaining good local and overall geometric quality of the structures [21].

2.2 Progress in restraints

The quality of a solution structure increases with the number of restraints used in the structure calculation [22]. The total number of restraints, sometimes reported as number of restraints-per-residue, is the most straightforward parameter to evaluate a structure

quality. The precision to which a structure can be determined is directly related to the number of experimental restraints used for the structure calculations. In case of data from ^{1}H NOE alone, about 15 restraints per residue can be typically obtained, which provide a family of conformers with backbone RMSD around 0.5 Å for well-folded regions. NOEs per residue can increase until 20-25 restraints-per-residues for double-labeled, ^{15}N and ^{13}C proteins thus leading to the much higher precision structures with backbone RMSD values in the range of 0.3 to 0.5 Å [22-24]. Regions affected by conformational disorder such as flexible fragments, large loops, terminal segments, are typically characterized by fewer restraints. Therefore, the number of restraints-per-residues decreases significantly when disordered regions are present.

The strong interest of the NMR community for new classes of restraints arises from the need of solving structures without - or with a limited numbers of - NOE distances. Indeed, the analysis of NOESY spectra can be complicated because of spectral overlap, spin diffusion and local dynamics. The restraints derived from NOE experiments are usually expressed in terms of allowed distance ranges. The lower limit is usually taken as the sum of two hydrogen atomic radii, i.e. 2 Å. The upper limits are derived from the NOE intensities, typically in the range of 2-5 Å. In folded proteins there are many proton pairs with an internuclear distance up to 5 Å, thus NOESY spectra tend to be very crowded and ambiguities in their assignments may be not resolved. Recently, to overcome this limitation, 4D NOESY spectra has been implemented using uniformly ^{13}C,^{15}N-labeled samples [25]. Although these experiments facilitate the assignment of NOEs in large proteins, they may suffer from a lack of resolution as the protein size increases. Another strategy for tacking large proteins is the replacement of most, if not all, hydrogen atoms in a protein with deuterium. In this way ^{13}C transverse relaxation rates are reduced, increasing resolution and allowing detection of cross peaks in scalar coupling based experiments which can lead to a complete spectral assignment. NOESY experiments performed on deuterated samples, where longer mixing times can be used, enable the detection of much longer range NOEs (up to 8 Å) between amide protons [26].

The consequences of spectral crowdness are two-fold: on the one hand the analysis of NOESY spectra and their conversion into distance restraints is typically the rate limiting step in a structure calculation; on the other hand NOESY data are more easily mis-

estimated and mis-interpreted than other restraints. Therefore, the study of new classes of structural restraints is essential in the framework of high-throughput structure determination aiming at the obtainment of fast and reliable information on the protein scaffold.

Initially, NOEs and ^{3}J coupling constant values, obtained from HNHA and HNHB experiments, were basically the most important structural restraints [27]. At the end of the last century, technological and methodological advancements made easy accessible the collection and analysis of restraints such as: residual dipolar couplings (RDCs), which are present when the protein is partially oriented in high magnetic fields in the presence of orienting media [28-35]; hydrogen bonds [36,37]; relative orientation of internuclear vectors as obtained from cross-correlation effects [38]; information on secondary structure derived from Chemical Shift Index [39,40]; backbone angle restraints from the analysis of a database for chemical shift and sequence homology [41]; a number of additional structural restraints (mostly pseudocontact shifts and nuclear relaxation enhancements, in addition to paramagnetic RDCs) derived from the presence of paramagnetic probes [42-44]. Furthermore, computational developments allowed users to well-implement these restraints into algorithms used for the structural calculations and to explore the potentiality of each type of restraints [27,45-48].

In general, NMR data alone would not be sufficient to determine the positions of all atoms in a biological macromolecule. It must be supplemented with information about the covalent structure of the protein like the amino acid sequence, bond lengths, bond angles, chiralities and planar groups as well as by steric repulsion between non-bonded atoms pairs.

2.2.1 Residual dipolar coupling

Inclusion of RDC restraints generally improves the precision of families of conformers [49]. A plot of experimental versus calculated RDC restraints for the structures prior to and after their inclusion in structure calculation, usually shows an increase in the agreement between the measured and calculated RDCs [50].

The measurement of residual dipolar coupling provides distance-independent structural restraints which, in combination with classical NOE data, have been shown to improve structure determination in multidomain systems and protein-protein or protein-ligand complexes [51-55]. Furthermore, residual dipolar couplings are powerful restraints, not only to complement NOE in structure determination [56] but also to determine an ab initio fold, even in the absence of NOE data [57,58]. In contrast to NOEs, interpretation of RDCs in terms of structure and dynamics is susceptible to relatively few sources of error and therefore they should provide more precise structural definition [59,60].

RDCs in the absence of NOEs have been used to calculate structural models by fitting RDCs to homologous structures [61]. RDCs alone have been also used to simultaneously determine the protein fold and the conformational dynamics [62,63]. In case of systems characterized by well defined secondary structures, such as a protein mainly consisting of α-helices, the relative orientation of the helices can often be established from RDCs, not only for water-soluble proteins but also for systems solubilized in detergents or embedded in lipid bilayers [64,65]. However, to date, no protein structures have yet been reported that are based exclusively on RDCs without NOEs or a model system.

A problem related with the use of dipolar couplings in structure determination is that a dipolar coupling does not uniquely describe an internuclear vector orientation. To determine directly the orientation of the internuclear vector, defined by the two dipole-dipole coupled nuclei in a arbitrary reference frame, RDCs for more than two noncollinear alignment conditions need to be measured [66].

RDCs originating from the presence of a paramagnetic metal ion were first used for solution structure determination in 1998, to refine the NMR structure of cytochrome b5 [56]. Their main advantage with respect to diamagnetic RDCs is that an independent and accurate estimate of the orientating tensor is available from pseudo-contact shift values [67,68]. Indeed, when external orienting systems are used, the orientation tensor parameters from the RDCs themselves may be underestimated as a result of the presence of local motions [69].

The ability of other than NOEs restraints, such as dihedral angles, hydrogen bonds, RDCs and paramagnetic-based restraints such as pseudocontact shifts, RDCs,

longitudinal relaxation rates, to improve the structure resolution and their complementarity with NOEs has been discussed by Bertini et al [70] for the calcium binding protein, Calbindin D_{9k}, in which one of the two calcium ions has been substituted by different lanthanide ions.

2.3 How to assess the quality of NMR structures

NMR structure quality can be assessed by comparing the starting experimental data with those resulting from the structures. Some peculiar features can be used such as: the number of total restraints (already discussed), violations of restraints and maximum consistent violations within the family structures, RMSD and target function values of the ensemble of conformers. Their analysis allows us to define the quality of a calculated family.

The most widely used method to describe the agreement of calculated structures with experimental data is the number, magnitude and consistency of violations. NMR structures are rarely in exact agreement with the whole experimental restraints used to calculated them [71]. For example, internal dynamics and other sources of line broadening, such as the presence of several different conformers in solution, can determine mis-calibration for some sets of NOEs and therefore can generate inconsistencies in the input data and produce violations during the calculations.

The quality of the structures is also determined by the number of violations, above an accepted threshold, present in a significant number of conformers. Presently, there is a general consensus that structures with violations up to 0.5 Å can be acceptable [72]. However, even if structures do not have violations above 0.5 Å they may still have a significant number of violations in the 0.3-0.4 Å range. Their relevance with respect to the structural quality is related to their consistency. Restraint violations are usually considered as consistent if the violation occurs in more than 75% of an ensemble of conformers [73]. The analysis of consistent restraint violations is an important aspect in the assessment of NMR structures because it points out at protein regions that need to be further analyzed in order to improve the agreement with the experimental data.

The agreement of the conformers in the family with the experimentally derived structural restraints can be judged by the RMSD between the experimental value of the restraint with the value measured on the calculated structures [74]. In general, the lower the target function, the better the fit of the structure to the experimental data. Recently, Montelione et al have proposed a new NMR quality assessment score (NMR RPF) which efficiently provides global measure of the goodness of the fit of the 3D structures to NOESY data, using methods from information retrieval statistics [75]. In general, the quality assessment of NMR structures can be performed by evaluating some parameters such as distance violations, restraints per residue and the RMSD, but not always they are correlated with a good RPF scores. Basically, is possible to obtain an ensemble of NMR structures with low RMSD but, at in the same time, without a reliable assessment of the accuracy of structures or how well the structures fit the experimental data. RPF scores directly measure the quality of structures against the NOESY peak list data and not against the restraints list used in the structural calculation. NOESY peak lists provide information that are closer to the raw NMR spectra than restraint lists, which might be affected by higher levels of mis-interpretation [75].

Usually, a good ensemble of conformers minimizes the violations of restraints and consequently the target function value [21]. In the assessment, it is important to evaluate the effect of introducing different types of restraints on the structure calculation. For this, their consistency with the restraints used in initial calculations and, consequently, their effects on the accuracy and precision of the structure should be evaluated. This can be determined by comparing the RMSD of the conformers in the family before and after introducing new restraints in the calculation. The new class of restraints is consistent with the other ones if no meaningful increase in violations and in the target function value are observed [70]. Therefore, the new restraints contributes to increase the precision of the structure if they produce a lower RMSD among of conformers.

The final step of NMR structures calculation is its energy minimization, with the inclusion of restraints, as pseudopotentials. Possibly equilibration of the protein conformation through some steps of Molecule Dynamic Calculations is also useful. Energy minimization can be performed using a box of water in which the molecule is placed. The most commonly used software are CHARMM [76], AMBER [77] and GROMOS [78].

Tools commonly used to evaluate different aspects of the quality protein structure, based on the knowledge of protein structural data, include PROCHECK [79,80], MolProbity [81], Verify3D [82], ProsaII [83], the PDB validation software [84] (now available under a single interface analyses called PSVS, Protein Structure Validation Software)[85], AQUA [79],WHAT IF [86] and QUEEN [87].

References

[1] G.N. Ramachandran, C. Ramakrishnan, V. Sasisekharan, J. Mol. Biol. 7 (1963), 95-99.

[2] I.C. Felli, B. Brutscher, ChemPhysChem 10 (2009), 1356-1368.

[3] L. Frydman, T. Scherf, A. Lupulescu, Proc. Natl. Acad. Sci. USA 99 (2002), 15858-15862.

[4] D. Marion, J. Biomol. NMR 32 (2005), 141-150.

[5] E. Kupce, R. Freeman, J. Am. Chem. Soc. 126 (2004), 6429-6440.

[6] B. Brutscher, J Biomol NMR 29 (2004), 57-64.

[7] E. Lescop, B. Brutscher, J. Am. Chem. Soc. 129 (2007), 11916-11917.

[8] D. Malmodin, M. Billeter, J. Am. Chem. Soc. 127 (2005), 13486-13487.

[9] D.K. Staykova, J. Fredriksson, W. Bermel, M. Billeter, J. Biomol. NMR 42 (2008), 87-97.

[10] E.L. Ulrich, H. Akutsu, J.F. Doreleijers, Y. Harano, Y.E. Ioannidis, M. Lin, S. Mading, D. Maziuk, Z. Miller, E. Nakatani, C.F. Schulte, D.E. Tolmie, R. Wenger, H. Yao, J.L. Markley, Nucl. Acid Res. 36 (2008).

[11] Y. Shen, O. Lange, F. Delaglio, Proc. Natl. Acad. Sci. U. S. A 105 (2008), 4685-4690.

[12] J. Kuszewski, A.M. Gronenborn, G.M. Clore, J. Am. Chem. Soc. 121 (1999), 2337-2338.

[13] J.P. Linge, M. Nilges, J Biomol NMR 13 (1999), 51-59.

[14] R. Sprangers, M.J. Bottomley, J.P. Linge, J. Schultz, M. Nilges, M. Sattler, J. Biomol. NMR 16 (2000), 47-58.

[15] G. Cornilescu, J. Marquardt, M. Ottiger, A. Bax, J. Am. Chem. Soc. 120 (1998), 6836-6837.

[16] W. Gronwald, R. Kirchhofer, A. Gorler, W. Kremer, B. Ganslmeir, K.P. Neidig, H.R. Kalbitzer, J Biomol NMR 17 (2000), 137-151.

[17] A.M. Gronenborn, G.M. Clore, CRC Crit. Rev. Biochem. Mol. Biol. 30 (1995), 351-385.

[18] A.M.J.J. Bonvin, R. Boelens, R. Kaptein, J. Biomol. NMR 1 (1991), 305.

[19] D. Zhao, O. Jardetzky, In: (239 ed.) (1994), pp. 601-607.

[20] G.M. Clore, M.A. Robien, A.M. Gronenborn, J Mol Biol 231 (1993), 82-102.

[21] C.A. Spronk, S.B. Nabuurs, A.M. Bonvin, E. Krieger, G.W. Vuister, G. Vriend, J Biomol NMR 25 (2003), 225-234.

[22] J.F. Doreleijers, M.L. Raves, T. Rullmann, R. Kaptein, J. Biomol. NMR 14 (1999), 123-132.

[23] J.F. Doreleijers, J.A.C. Rullmann, R. Kaptein, J. Mol. Biol. 281 (1998), 149-164.

[24] J. Cavanagh, W.J. Fairbrother, A.G. Palmer, III, M. Rance, N.J. Skelton, Protein NMR Spectroscopy. Principles and practice, Academic Press, SanDiego (2007).

[25] D.A. Snyder, Y. Xu, D. Yang, R. Bruschweiler, J Am Chem Soc. 129 (2007), 14126-14127.

[26] L.M.I. Koharudin, A.M.J.J. Bonvin, R. Kaptein, R. Boelens, J. Magn. Reson. 163 (2003), 228-235.

[27] P. Guntert, W. Braun, K. Wüthrich, J. Mol. Biol. 217 (1991), 517-530.

[28] N. Tjandra, A. Bax, Science 278 (1997), 1111-1114.

[29] J.R. Tolman, J.M. Flanagan, M.A. Kennedy, J.H. Prestegard, Proc. Natl. Acad. Sci. USA 92 (1995), 9279-9283.

[30] N. Tjandra, S. Grzesiek, A. Bax, J. Am. Chem. Soc. 118 (1996), 6264-6272.

[31] N. Tjandra, J.G. Omichinski, A.M. Gronenborn, G.M. Clore, A. Bax, Nature Struct. Biol. 4 (1997), 732-738.

[32] R.R. Vold, P.S. Prosser, J. Magn. Reson. Ser. B 113 (1996), 267-271.

[33] T. Carlomagno, W. Peti, C. Griesinger, J. Biomol. NMR 17 (2000), 99-109.

[34] G. Otting, M. Ruckert, M.H. Levitt, A. Moshref, J. Biomol. NMR 16 (2000), 343-346.

[35] W. Peti, C. Griesinger, J. Am. Chem. Soc. 122 (2000), 3975-3976.

[36] A.J. Dingley, S. Grzesiek, J. Am. Chem. Soc. 120 (1998), 8293-8297.

[37] F. Cordier, S. Grzesiek, J. Am. Chem. Soc. 121 (1999), 1601-1602.

[38] M. Rief, M. Gautel, F. Oesterhelt, J.M. Fernandez, H.E. Gaub, Science 276 (1997), 1109-1112.

[39] D.S. Wishart, B.D. Sykes, F.M. Richards, J. Mol. Biol. 222 (1991), 311-333.

[40] S.M. Gagne', S. Tsuda, M.X. Li, M. Chandra, L.B. Smillie, B.D. Sykes, Protein Sci. 3 (1994), 1961-1974.

[41] G. Cornilescu, F. Delaglio, A. Bax, J Biomol NMR 13 (1999), 289-302.

[42] I. Bertini, C. Luchinat, M. Piccioli, Methods Enzymol. 339 (2001), 314-340.

[43] R. Barbieri, C. Luchinat, G. Parigi, ChemPhysChem 21 (2004), 797-806.

[44] G.M. Clore, A.M. Gronenborn, A. Bax, J. Magn. Reson. 133 (1998), 216-221.

[45] P. Guntert, C. Mumenthaler, K. Wüthrich, J. Mol. Biol. 273 (1997), 283-298.

[46] L. Banci, I. Bertini, M.A. Cremonini, G. Gori Savellini, C. Luchinat, K. Wüthrich, P. Güntert, J. Biomol. NMR 12 (1998), 553-557.

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

[48] C.D. Schwieters, J. Kuszewski, N. Tjandra, G.M. Clore, J. Magn. Reson. 160 (2003), 65-73.

[49] R.S. Lipsitz, N. Tjandra, Annu Rev Biophys Biomol Struct 33 (2004), 387-413.

[50] J.R. Tolman, J. Am. Chem. Soc. 124 (2002), 12020-12030.

[51] N. Tjandra, Structure Fold Des. 7 (1999), R205-R211.

[52] M.W. Fischer, J.A. Losonczi, J.L. Weaver, J.H. Prestegard, Biochemistry 38 (1999), 9013-9022.

[53] G.M. Clore, M.R. Starich, M. Bewley, M. Cai, J. Kuszewski, J. Am. Chem. Soc. 121 (1999), 6513-6514.

[54] P.J. Bolon, H.M. Al-Hashimi, J.H. Prestegard, J. Mol. Biol. 293 (1999), 107-115.

[55] M. Cai, H. Wang, E.T. Olejniczak, R.P. Meadows, Gunasekera A.H., Xu N., S.W. Fesik, J. Magn. Reson. 139 (1999), 451-453.

[56] L. Banci, I. Bertini, J.G. Huber, C. Luchinat, A. Rosato, J. Am. Chem. Soc. 120 (1998), 12903-12909.

[57] F. Delaglio, G. Kontaxis, A. Bax, J. Am. Chem. Soc. 122 (2000), 2142-2143.

[58] J.C. Hus, D. Marion, M. Blackledge, J. Am. Chem. Soc. 123 (2001), 1541-1542.

[59] J.H. Prestegard, H.M. Al-Hashimi, J.R. Tolman, Q Rev Biophys 33 (2000), 371-424.

[60] M. Blackledge, Prog. NMR Spectrosc. 46 (2005), 23-61.

[61] J.J. Chou, S. Li, C.B. Klee, A. Bax, Nature Struct. Biol. 8 (2001), 990-997.

[62] G. Bouvignies, S. Meier, S. Grzesiek, M. Blackledge, Angew. Chem Int. Ed Engl. 45 (2006), 8166-8169.

[63] G. Bouvignies, P. Markwick, R. Brüschweiler, M. Blackledge, J Am Chem Soc. 128 (2006), 15100-15101.

[64] A.A. De Angelis, D.H. Jones, C.V. Grant, S.H. Park, M.F. Mesleh, S.J. Opella, Nuclear Magnetic Resonance of Biologica Macromolecules, Pt C 394 (2005), 350-382.

[65] S.C. Howell, M.F. Mesleh, S.J. Opella, Biochemistry 44 (2005), 5196-51206.

[66] A. Bax, Protein Sci 12 (2003), 1-16.

[67] I. Bertini, C. Luchinat, G. Parigi, Prog. NMR Spectrosc. 40 (2002), 249-273.

[68] I. Bertini, A. Bhaumik, G. De Paepe, R.G. Griffin, M. Lelli, J.R. Lewandowski, C. Luchinat, In: (2009).

[69] I. Bertini, C. Luchinat, G. Parigi, R. Pierattelli, ChemBioChem 6 (2005), 1536-1549.

[70] I. Bertini, A. Donaire, B. Jiménez, C. Luchinat, G. Parigi, M. Piccioli, L. Poggi, J. Biomol. NMR 21 (2001), 85-98.

[71] Sander B.N., C.A.E.M. Spronk, G. Vriend, G.W. Vuister, Concepts Magn. Reson. 22A (2004), 90-105.

[72] C.A.E.M. Spronk, S.B. Nabuurs, E. Krieger, G. Vriend, G.W. Vuister, Prog. NMR Spectrosc. 45 (2004), 315-337.

[73] P. Guntert, Q. Rev. Biophys. 31 (1998), 145-237.

[74] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne, Nucl. Acids Res. 28 (2000), 235-242.

[75] Y.J. Huang, R. Powers, G.T. Montelione, J Am Chem Soc. 127 (2005), 1665-1674.

[76] S.R. Prakash P, In: (2009).

[77] D.A. Case, T.A. Darden, T.E. Cheatham, C.L. Simmerling, J. Wang, R.E. Duke, R. Luo, K.M. Merz, B. Wang, D.A. Pearlman, M. Crowley, S. Brozell, V. Tsui, H. Gohlke, J. Mongan, V. Hornak, G. Cui, P. Beroza, C.E. Schafmeister, J.W. Caldwell, W.S. Ross, P.A. Kollman, In: University of California, San Francisco, CA (2008).

[78] M. Christen, P.H. Hunenberger, D. Bakowies, R. Baron, R. Burgi, D.P. Geerke, T.N. Heinz, M.A. Kastenholz, V. Krautler, C. Oostenbrink, C. Peter, D. Trzesniak, W.F. Van Gunsteren, J Comp Chem 26 (2005), 1719-1751.

[79] R.A. Laskowski, J.A.C. Rullmann, M.W. MacArthur, R. Kaptein, J.M. Thornton, J. Biomol. NMR 8 (1996), 477-486.

[80] R.A. Laskowski, M.W. MacArthur, D.S. Moss, J.M. Thornton, J. Appl. Crystallogr. 26 (1993), 283-291.

[81] S.C. Lovell, I.W. Davis, W.B. Arendall, III, P.I. de Bekker, J.M. Word, M.G. Prisant, J.S. Richardson, D.C. Richardson, Proteins 50 (2003), 437-450.

[82] R. Luthy, J.U. Bowie, D. Eisenberg, Nature 356 (1992), 83-85.

[83] M.J. Sippl, Proteins Struct. Funct. Genet. 17 (1993), 355-362.

[84] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne, Nucleic Acids Res. 28 (2000), 235-242.

[85] A. Bhattacharya, R. Tejero, G.T. Montelione, Proteins 66 (2007), 778-795.

[86] G. Vriend, J. Mol. Graphics 8 (1990), 52-56.

[87] S.B. Nabuurs, C.A. Spronk, E. Krieger, H. Maassen, G. Vriend, G.W. Vuister, J. Am. Chem. Soc. 125 (2003), 12026-12034.

Chapter 3

3. NMR characterization of the slow dynamics in biomacromolecules

The flexibility of a protein is closely coupled to its function, as observed from many atomic-resolution structures that show large-scale conformational changes between different states, such as in the presence or absence of a bound ligand.

Protein function relies on structural protein dynamics, with time scales ranging from picoseconds to beyond seconds. In the case of molecular recognition, for example, proteins adapt their structure to different binding partners, often exhibiting large structural heterogeneity. In the past 30 years, information, at atomic scale resolution, on many dynamical processes has been accumulated from NMR spectroscopy and X ray crystallography.

NMR spectroscopy can be used to monitor the dynamic behavior of a protein at atomic resolution. Moreover, protein motions over a broad range of timescales can be monitored using various types of NMR experiments (Fig. 1): nuclear spin relaxation rate measurements report the internal motions on fast (subnanoseconds) and slow (microseconds to milliseconds) timescales as well as the overall rotational diffusion of the molecule (5–50 nanoseconds), whereas rates of magnetization transfer among protons with different chemical shifts and proton exchange report movements of protein domains on slower timescales (from milliseconds to days). These features make NMR a unique and powerful tool in studying protein dynamics related to protein functions, and there has been a tremendous growth in these applications [1]. In parallel with new applications, there have been important developments in experimental methods that significantly increased the accuracy of the information obtained, and thus expanded the range of questions that can be addressed. The basic theory and experimental approaches underlying the current advances were established about 30–40 years ago [2]. Over the past 20 years, the intense development of heteronuclear NMR spectroscopy of biomolecules isotopically enriched with ^{2}H, ^{13}C and/or ^{15}N has made studies of dynamics at atomic resolution possible.

In this work, we focus the attention only on the characterization of the slow dynamics on the microseconds to milliseconds time scales.

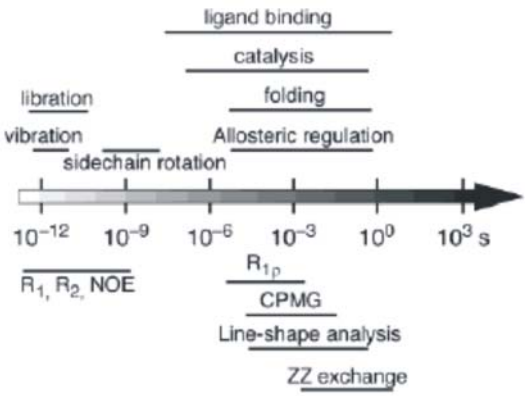

Figure 1. NMR time scales for protein dynamics [3].

3.1 NMR relaxation and molecular dynamics

Relaxation is the process that restores a spin system to its equilibrium. In solution NMR, relaxation is mediated by random processes that interact with the spin system. These are typically due to molecular reorientation, internal motions or chemical exchange involving the nuclei of interest [1,3-8]. A first consequence of motion is that anisotropic spin–spin interactions are averaged out and that the NMR spectra shows only the effect of isotropic interactions. The time dependence of anisotropic interactions does however contribute to relaxation, and is often the dominant effect. The time dependent magnetic field fluctuations are generated by two principal mechanisms, the anisotropic chemical shift (chemical shift anisotropy, CSA) associated with a single spin and the dipolar coupling between nearby spins. Both of these mechanism operate simultaneously for all types of spins.

The external B_o field is reduced at the nucleus by the surrounding electron density, by a shielding factor σ, giving an observed chemical shift of $\omega s = \gamma(1 - \sigma)B_o$. If the electron density is the same in all directions (isotropic) then a change in the orientation of the spin will have no effect on the shielding of the nucleus, and therefore no effect on the magnetic field at the nucleus. If the shielding is anisotropic, then different orientations due to molecular motion will generate different magnetic fields at the nucleus. These varying magnetic fields are a source of relaxation. The actual field will depend on the orientation of the molecule with respect to the external magnetic field. To characterize

the anisotropic nature of the shielding we will define a chemical shift tensor, σ^*, which will give the chemical shift for any given orientation of the molecule with respect to the B_o field. The chemical shift tensor is related to the shielding as follows: $\delta^* = (1 - \sigma^*)$, where σ^* is the shielding tensor. The approaches to determine the CSA of the nuclear spins are beyond the scope of this text, details can be found in literature [9].

Any nucleus with a non-zero spin angular moment generates an instantaneous magnetic dipolar field that is proportional to the magnetic moment of the nucleus. In presence of molecular motions in solution, this field can fluctuate and constitutes a mechanism for relaxation of nearby spins.

Dipolar coupling arises when the magnetic field of one nuclear spin affects the local magnetic field of another spin. The intensity of the dipole field, B_d , depends on both the orientation of the two spins and the distance between them (inverse sixth power of the distance). If either of these properties are time dependent then the magnetic field will vary with time.

Another phenomena that can contribute to spin relaxation of the nuclei are the relaxation interference between different relaxation mechanisms. Correlations between two stochastic Hamiltonians results in cross-correlated relaxation, if both Hamiltonians are affected by the same molecular motions. For example, the interference between the dipolar and chemical shift anisotropy relaxation mechanism are known since many years [10] and the effect of this phenomena is that the two lines in a scalar coupled doublet have a different linewidths.

In general, the occurrence of interference between two relaxation mechanisms is such that the cross correlated relaxation can be added or subtracted to the auto-relaxation, thus giving rise to different relaxation properties for the same coherence. This, in turn, can be a source of coherence transfer (also termed: Relaxation Allowed Coherence Ttransfer).

The return to equilibrium of a spin system can be described from a macroscopic point of view by relaxation rates, which account for the decay of the detected signal as a function of time. Relaxation rates for nuclei in proteins depend upon a large number of factors, including: overall rotational correlation times, internal motions, the geometrical arrangement of nuclei, and the relative strengths of the applicable relaxation mechanisms. If the overall correlation time and the three-dimensional structural coordinates of the protein are known, relaxation rates can be calculated in a relatively

straightforward manner using expressions derived elsewhere [11]. In general, ^{1}H relaxation in proteins is dominated by dipolar interactions with other protons (within approximately 5Å) and by interactions with directly bonded heteronuclei. The latter arise from dipolar interactions with ^{13}C and ^{15}N in labeled proteins or from scalar relaxation of the second kind between the quadrupolar ^{14}N nuclei and amide protons. Relaxation of protonated ^{13}C and ^{15}N heteronuclei is dominated by dipolar interactions with the directly bonded protons, and secondarily by CSA (for ^{15}N spins and aromatic ^{13}C spins). Relaxation of unprotonated heteronuclei, carbonyl ^{13}C and unprotonated aromatic ^{13}C spins, is dominated by CSA interactions.

3.2 Relaxation techniques for slow dynamics: Chemical Shift Modulation

Chemical exchange is a ubiquitous phenomenon in NMR spectroscopy that provides information on conformational and chemical kinetic process occurring on μs-s time scales. Chemical exchange is the manifestation of processes that modulate isotropic chemical shifts by altering the magnetic environments of spins. In this case the chemical shift changes may be large and the NMR signal broad, leading to an increase in R_2. The main experimental techniques for quantifying chemical exchange are longitudinal magnetization exchange [12], line shape analysis [13], CPMG relaxation dispersion [14] and $R_{1\rho}$ relaxation dispersion [14]. These techniques are most commonly applied to ^{1}H, ^{13}C, ^{15}N, and ^{31}P spins in biological macromolecules.

Recent spectroscopic developments have focused on extracting the exchange contribution in multiple quantum relaxation rates (ΔR^{MQ}). If two spins are affected by the same chemical exchange kinetic process, the chemical shift changes for the two spins resulting from transitions between two sites are correlated. This correlation can either broaden or narrow resonance lineshapes for multiple quantum coherences [15].

Local motions such as rotations about dihedral angles affect both anisotropic and isotropic components of the chemical shifts. Motions that are slower than the correlation time τ_c of the molecule can lead to a modulation of the isotropic shifts of neighboring nuclei, and thus contribute to the relaxation of coherences [16,17]. By contrast, the anisotropic components (CSAs) are averaged out by overall tumbling in the intervals between rare conformational rearrangements. Multiple quantum coherences (MQCs) will not only be sensitive to the fluctuations of the chemical shifts of the nuclei involved

in the coherences, but also to cross-correlation between the modulations of their chemical shifts (CSM/CSM). Since these interferences affect zero quantum coherences (ZQCs) and double quantum coherences (DQCs) in a different manner, the difference between their relaxation rates allows one to determine the cross-correlation rates [18].
Indeed, correlated isotropic chemical shift modulations (CSM/CSM) is not the only mechanism contributing to the difference ΔR between the relaxation rates of zero- (ZQ) and double-quantum (DQ) coherences [16], as also two dipole-dipole interactions (DD/DD) and two chemical shift anisotropy interactions (CSA/CSA), contribute to correlated fluctuations (Fig. 2).

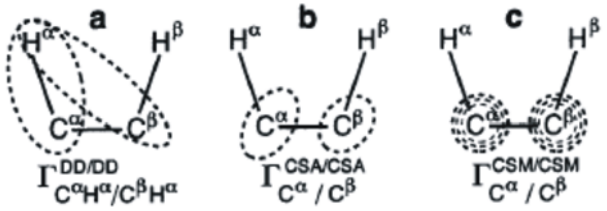

Figure 2. Selected cross-correlated mechanisms that affect the relaxation of $C^\alpha C^\beta$ multiple quantum coherences. (a-c) Cross-correlated effects that contribute to auto-relaxation. (a) Dipole/dipole cross-correlation involving a single external spin (here H^α), (b) CSA/CSA cross-correlation, and (c) correlated isotropic chemical shift modulations (CSM). These mechanisms induce differential line broadening when comparing the ZQ and DQ line widths [16].

The differential broadening ΔR^{MQ} can be decomposed into three terms:

$$\Delta R^{MQ} = R^{CSM/CSM} + R^{DD/DD} + R^{CSA/CSA} \quad (1)$$

The $R^{CSM/CSM}$ contribution accounts for fluctuations of isotropic chemical shifts that affect simultaneously the environments of the nuclei involved into MQCs. They are characterized by time-scales that typically fall in the µs to ms range [19-21]. These chemical shift fluctuations can be caused by changes of dihedral φ and ϕ backbone angles, as can be rationalized by *ab initio* calculations [22]. In the case of metal-binding proteins, they can also result from interactions with metal ions. The sign of the $R^{CSM/CSM}$

contribution depends on the signs of the changes of the chemical shifts upon conformational exchange. The modulations of the isotropic shifts are said to be *correlated* or *anti-correlated* if the $R^{CSM/CSM}$ rates are positive or negative.

The second term in Eq. (1) comprises contributions from various auto- and cross-correlated dipole-dipole effects [17,23]. As discussed elsewhere [15,18] the contributions to $R^{DD/DD}$ due to various dipolar auto-relaxation processes can safely be neglected, and cross-relaxation processes are estimated from the 3D structure. The third term in Eq. (1) arises from cross-correlated CSA relaxation of the nuclei due to concerted modulations of the chemical shifts that are caused by molecular motions.

Using this innovative approach, in this work we have studied the μs-ms fluctuations of the peptide plane in proteins by direct detection of C'N multiple quantum relaxation rates.

References

[1] L.E. Kay, Nature Struct. Biol. 5 (1998), 513-517.

[2] J.I. Kaplan, G. Fraenkel, In: NMR of Chemically Exchanging Systems, Academic Press, New York (1980).

[3] A.G. Palmer, III, Chem. Rev. 104 (2004), 3623-3640.

[4] V. Daragan, K.H. Mayo, Progr. NMR Spectrosc. 31 (1997), 63-105.

[5] B. Brutscher, Concepts Magn. Reson. 12 (2000), 207-229.

[6] R. Brüschweiler, D.A. Case, Phys. Rev. Lett. 72 (1994), 940-943.

[7] M.W.F. Fischer, L. Zeng, A. Majumdar, E.R.P. Zuiderweg, Proc. Natl. Acad. Sci. USA 95 (1998), 8016-8019.

[8] Kumar A., R.C.R. Grace, P.K. Madhu, Prog. NMR Spectrosc. 37 (2000), 191-319.

[9] J. Cavanagh, W.J. Fairbrother, A.G. Palmer, III, M. Rance, N.J. Skelton, Protein NMR Spectroscopy. Principles and practice, Academic Press, SanDiego (2007).

[10] M. Goldman, J. Magn. Reson. 60 (1984), 437-452.

[11] A.G. Palmer, III, Curr Opin Struct Biol 7 (1997), 732-737.

[12] J.J. Led, H. Gesmar, F. Abilgaard, In: (176 ed.) (1989), pp. 311-329.

[13] J. Sandstrom, Dynamic NMR Spectroscopy, Academic Press, London (1982).

[14] D.G. Davis, M.E. Perlman, R.E. London, J. Magn. Reson. Ser. B 104 (1994), 266-275.

[15] D. Frueh, Prog. NMR Spectrosc. 41 (2002), 305-324.

[16] D. Frueh, J.R. Tolman, G. Bodenhausen, C. Zwahlen, J Am Chem Soc. 123 (2001), 4810-4816.

[17] R. Konrat, H. Sterk, Chem Phys Lett. 203 (1993), 75.

[18] F. Kateb, D. Abergel, Y. Blouquit, P. Duchambon, C.T. Craescu, G. Bodenhausen, Biochemistry 45 (2006), 15011-15019.

[19] C. Perazzolo, J. Wist, K. Loth, L. Poggi, S.W. Homans, G. Bodenhausen, J. Biomol. NMR 33 (2005), 233-242.

[20] J. Wist, C. Perazzolo, G. Bodenhausen, Appl. Magn. Reson. 29 (2005), 251-259.

[21] J. Wist, F. Dominique, J.R. Tolman, G. Bodenhausen, J Biomol NMR (2004), 263-272.

[22] W.D. Arnold, E. Oldfield, J Am Chem Soc. 122 (2000), 12835-12841.

[23] Pellecchia M., Pang Y., L. Wang, Kurochkin A.V., Kumar A., E.R.P. Zuiderweg, J. Am. Chem. Soc. 121 (1999), 9165-9170.

Chapter 4

4. Overview of the investigated proteins

4.1 Copper, Zinc Superoxide dismutase

Reactive oxygen species (ROS) are substances that are released during oxidative metabolism. ROS include the superoxide anion (O^{2-}), hydrogen peroxide (H_2O_2), and the hydroxyl radical ($OH^{\cdot}$) [1]. The superoxide anion is produced in aerobic metabolism following the one electron reduction of molecular oxygen occurring in respiration and photosynthesis [2,3] or during the immune response [4-6]. The reactions of ROS with macromolecules can lead to DNA mutations, changes in the structure and function of proteins, and peroxidative damage of cell membrane lipids [7]. Abundant evidence exists that ROS play an important role in the pathogenesis diseases [8].

Cells have an enzymatic antioxidant pathway against ROS which are generated during oxidative metabolism: superoxide dismutase (SOD) catalyzes the formation of hydrogen peroxide from superoxide radicals. Hydrogen peroxide can generate toxic hydroxyl radicals, but it is removed by a reaction catalyzed by catalase (CAT) and glutathione peroxidase (GPx) [9].

SOD enzymes are a class of metallo-enzymes which can bound Fe^{2+} (FeSOD), Mn^{2+} (MnSOD) and Cu^{2+},Zn^{2+} (CuZnSOD).

Several CuZnSODs from eukaryotic and prokaryotic sources have been characterized either by X-ray crystallography or NMR spectroscopy [10]. All these enzymes derive from a common ancestor gene. They display a similar three-dimensional fold, based on a flattened Greek-key eight-stranded β-barrel, and a conserved organization of the active site metal cluster, where a copper ion constitutes the catalytic center and a zinc ion plays mainly a structural role.

From the point of view of protein structure, the overall structures of the reduced and oxidized enzymes are extremely similar; the few big changes are localized to a very small region close to the Cu ion, which goes from three coordinate in Cu^{I}ZnSOD to five coordinate in Cu^{II}ZnSOD. In Cu^{I}ZnSOD, Cu(I) is bound by three histidyl imidazoles in a nearly trigonal-planar arrangement while Zn(II) is bound by three histidyl imidazoles and an Asp carboxylate. In Cu^{II}ZnSOD, the structure of most of the protein, as well as

the zinc site and the positions of the three His ligands to Cu, are changed very little from those in the reduced enzyme. The fundamental structural changes near the copper are two: one of the Zn ligands tilts, deprotonates, and binds to the Cu(II) as a bridging ligand and a water ligand is added to the first coordination sphere of Cu(II).

One particularly important structural feature that is identical in both the oxidized and reduced enzymes is the channel that provides access to the Cu ion for the substrate.

A mechanistic model for the catalytic reaction of CuZnSOD with superoxide is summarized in Fig. 3.

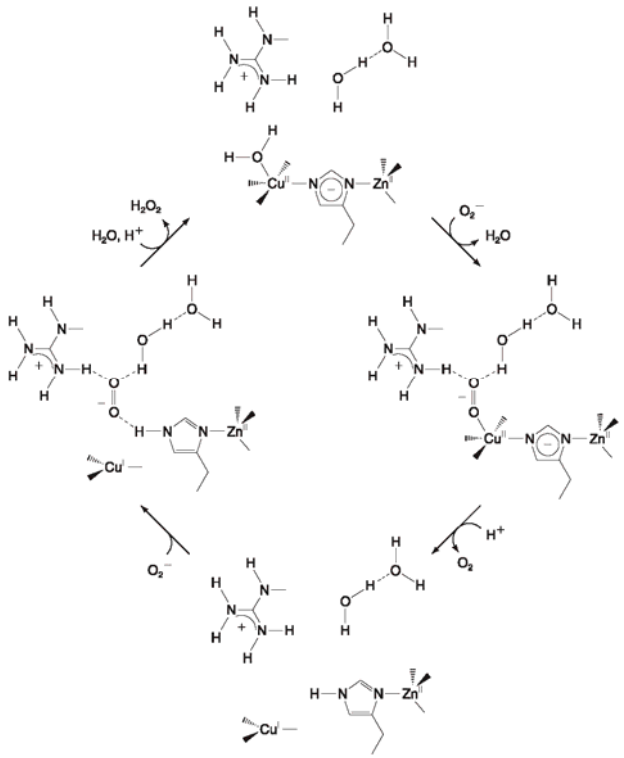

Figure 3. A schematic summary of the catalytic cycle of Cu,Zn superoxide dismutase [11].

In this work we studied the CuZnSOD from *Salmonella enterica*, a monomeric protein of 154 aminoacids. Protein overexpression was obtained following the protocol available in literature [12]. The 3D structure of this enzyme (*Sodc2*) (PDB code 2K4W) is shown in Fig. 4.

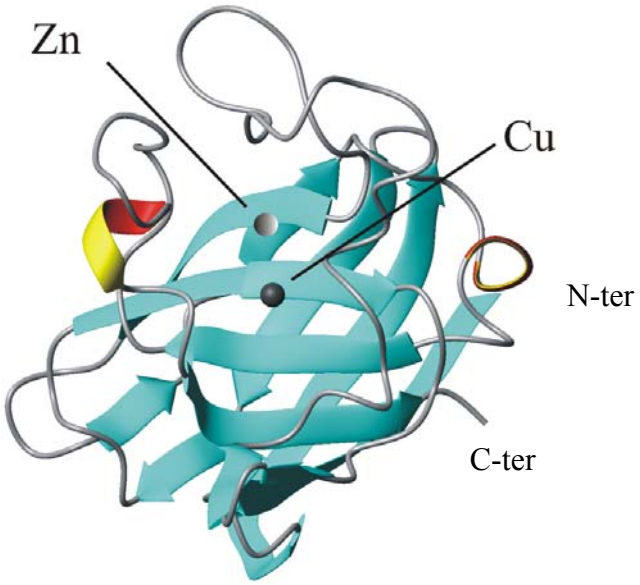

Figure 4. 3D structure of the enzyme *Sodc2* [12].

4.2 Matrix Metalloproteinases

Matrix metalloproteinases (MMP) are a class of extracellular zinc-containing peptidases that are involved in a variety of tissue-remodeling activities, and whose misfunction/misregulation is implicated in several pathologies ranging from arthritis rheumatoides to metastatic processes [13-17]. Such proteins, which may be bound to the outer cell membrane or released into the matrix as soluble molecules, are composed of a catalytic domain, and a hemopexin-like domain that probably plays a role in substrate recognition. In several pathologies overexpression of MMPs, or the misregulation of their activity, is related to disease progression [17]. MMPs thus are validated pharmaceutical targets. Many efforts have been devoted to develop inhibitors against these metalloenzymes [18] through drug design approaches, which require a detailed knowledge of the structural and dynamics features of the system.

In this work, the catalytic domain of Matrix Metalloproteinase 12 (MMP12), corresponding to the segment Gly-106–Gly-263, has been overexpressed as already reported in literature [19]. The catalytic domain was inhibited by NNGH (*N*-isobutyl-*N*-[4-methoxyphenylsulphonyl] glycyl hydroxamic acid). The structure of MMP12 (PDB code 1YCM) is shown in Fig. 5.

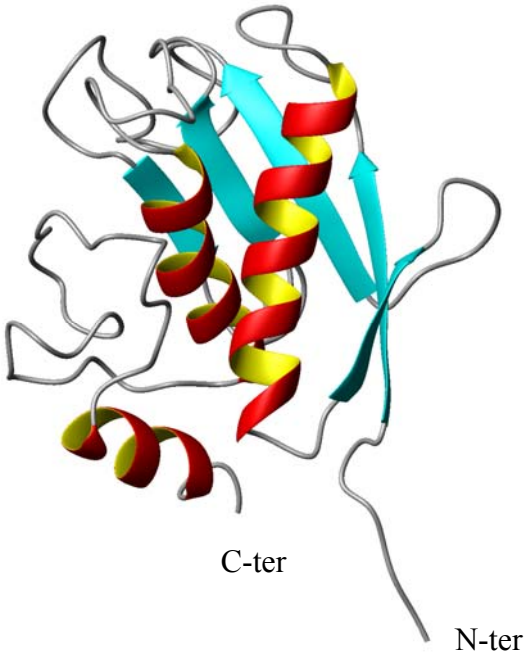

Figure 5. 3D structure of the MMP12 [19].

4.3 Calbindin D_{9k}

Calbindin D_{9k} is a protein of 75 residues that belongs to the S-100 subgroup of the calmodulin superfamily of intracellular calcium-binding proteins [20]. This family of proteins exhibits a wide diversity of functions ranging from maintaining intracellular Ca^{2+} levels to mediating specific cellular responses. Calbindin D_{9k} is involved in intracellular buffering of calcium ions and/or uptake of calcium ions from the intestinal brush border membrane and transport to the inner membrane [21]. Calbindin D_{9k} contains a single pair of EF-hands with the characteristic topology of the S-100 proteins: the N-terminal EF-hand has a variant ion-binding loop with 14 residues, as opposed to the 12 residues of a typical Calmodulin-like EF-hand. The protein binds two calcium ions with high affinity and positive cooperativity.

In this work protein expression and purification of native Ca^{2+} loaded Calbindin D_{9k} (Ca_2Cb) was performed as described elsewhere [22]. Lanthanide- (Ln^{3+}) substituted derivatives (CaCeCb) were obtained following the established procedure [23]. The 3D structure of Calbindin D_{9k} is shown in Fig. 6 (PDB code 1KSM), showing the organization of the four helices , the N-terminal and C-terminal Ca^{2+}-binding loops (loop I and loop II, respectively), and the linker region. The calcium ions in loop II can be easily replaced by a metal ion of the lanthanide series, here Ce^{3+} [24]. Such

substitutions lead to systems with similar structures but different spectroscopic properties, depending on the lanthanide ion.

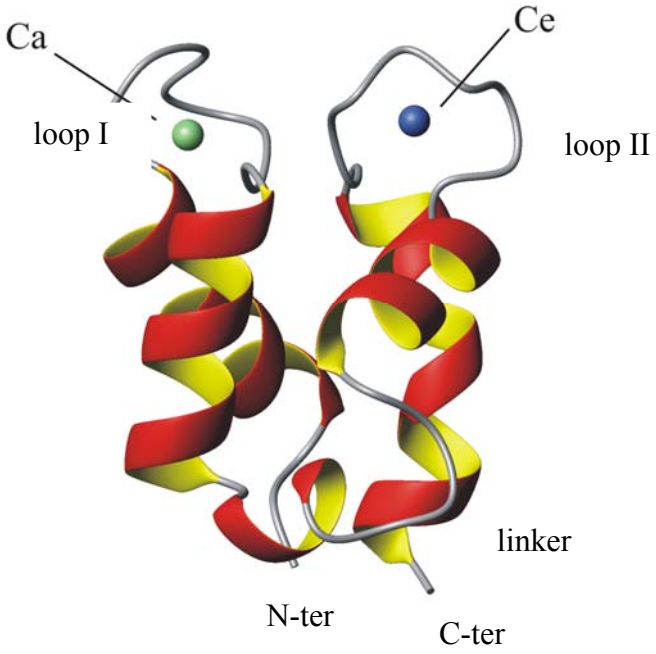

Figure 6. 3D structure of Calbindin D_{9k} where a native calcium ion is replaced with the lanthanide ion, Ce^{3+} [24].

References

[1] B. Halliwell, J.M. Gutteridge, Biochem. J. 219 (1984), 1-14.

[2] K.J. Davies, Biochem. Soc. Symp. 61 (1995), 1-31.

[3] C. Richter, V. Gogvadze, R. Laffranchi, R. Schlapbach, M. Schweizer, M. Suter, P. Walter, M. Yaffee, Biochim. Biophys. Acta 1271 (1995), 67-74.

[4] J.S. Kroll, P.R. Langford, B.M. Loynds, J Bacteriol 173 (1991), 7449-7457.

[5] J.D. Balentine, In: Pathology of Oxygen Toxicity, Academic Press, New York (1982).

[6] B. Meier, H.H. Radeke, S. Selle, M. Younes, H. Sies, K. Resch, G.G. Habermehl, Biochem. J. 263 (1989), 539-545.

[7] B. Halliwell, J.M.C. Gutteridge, D. Blake, Phil. Trans. Roy. Soc. ser. B 311 (1985), 659-671.

[8] Fang F.C., M.A. DeGroote, J.W. Foster, A.J. Bäumler, U. Ochsner, T. Testerman, S. Bearson, J.C. Giàrd, Y. Xu, G. CAmpbell, T. Laessig, Proc. Natl. Acad. Sci. USA 96 (1999), 7502-7507.

[9] C. Michiels, M. Raes, O. Toussaint, J. Remacle, In: (17 ed.) (1994), pp. 235-248.

[10] D. Bordo, A. Pesce, M. Bolognesi, M.E. Stroppolo, M. Falconi, A. Desideri, In: John Wiley and Sons, Editor, Handbook of Metalloproteins (Eds.Messerschmidt ed.), **Chichester.** (2001), pp. 1284-1300.

[11] L. Banci, I. Bertini, C. Luchinat, P. Turano, In: I. Bertini, B.H. Gray, J.S. Valentine, E.I. Stiefel, Editors, Biological Inorganic Chemistry: Structure & Reactivity, University Science Books, Sausalito (CA) (2006), pp. 229-261.

[12] M. Mori, B. Jiménez, M. Piccioli, A. Battistoni, M. Sette, Biochemistry 47 (2008), 12954-12963.

[13] H. Birkedal-Hansen, Curr. Opin. Cell Biol. 7 (1995), 728-735.

[14] G.J. Murphy, G. Murphy, J.J. Reynolds, FEBS Lett 89 (1991), 4-7.

[15] W.G. Stetler-Stevenson, R. Hewitt, Sem. Cancer Biol. 7 (1996), 147-154.

[16] N. Borkakoti, J Mol Med 78 (2000), 261-268.

[17] L. Steinman, Cell 85 (1996), 299-302.

[18] M. Whittaker, C.D. Floyd, P. Brown, A.J. Gearing, Chem. Rev. 99 (1999), 2735-2776.

[19] I. Bertini, V. Calderone, M. Cosenza, M. Fragai, Y.-M. Lee, C. Luchinat, S. Mangani, B. Terni, P. Turano, Proc. Natl. Acad. Sci. USA 102 (2005), 5334-5339.

[20] R.H. Kretsinger, CRC Crit. Rev. Biochem. 8 (1980), 119-174.

[21] C.W. Heizmann, W. Hunziker, Trends Biochem. Sci. 16 (1991), 98-103.

[22] W.J. Chazin, J. Kördel, T. Drakenberg, E. Thulin, P. Brodin, T. Grundstrom, S. Forsén, Proc. Natl. Acad. Sci. USA 86 (1989), 2195-2198.

[23] I. Bertini, C.J. Carrano, C. Luchinat, M. Piccioli, L. Poggi, Biochemistry 41 (2002), 5104-5111.

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

Chapter 5

5. The Solution Structure of the Monomeric Copper, Zinc Superoxide Dismutase from *Salmonella enterica*: Structural Insights To Understand the Evolution toward the Dimeric Structure

Superoxide dismutases (SOD) are a class of metalloenzymes which play a central role in the physiological response to oxygen toxicity. They catalyze the dismutation of the superoxide anion into molecular oxygen and hydrogen peroxide. Three different SODs are known, which have different catalytic centers containing, respectively: Fe^{2+} (FeSOD), Mn^{2+} (MnSOD) and $Cu^{2+}Zn^{2+}$ (CuZn SOD).

FeSOD is usually isolated as a dimer of identical subunits from prokaryotes and anaerobic bacteria, whereas MnSOD has been found as an oligomer of identical subunits in prokaryotes and in mitochondria of eukaryotic cells [1]. CuZnSODs have been reported in the vegetable world and in the animal kingdom, from invertebrate species to fish, reptiles, birds and mammals[2]. In eukaryotes, this enzyme is localized in the cytosol, in the intermembrane space and in lysosomes, while a distinct CuZnSOD is found as an extracellular form. CuZnSOD are also known to be widely distributed in the prokaryotic phyla . X-ray crystallographic studies have shown that the FeSOD and the MnSOD adopt very similar folds, but are unrelated to overall structure of the CuZnSOD [3].

In this thesis, we will focus our attention on CuZnSOD enzymes.

All eukaryotic CuZnSODs characterized up to now are composed of two identical subunits of about 2x16 kDa, structurally organized in an eight-stranded antiparallel β-barrel with Greek key topology, each carrying a Cu^{2+}, Zn^{2+} ion pair. The copper ion is the catalytic site of enzyme, whose physiological function takes place via a two step mechanisms:

$$Cu^{2+} + O_2^- \rightarrow Cu^+ + O_2$$

$$Cu^+ + O_2^- + 2H^+ \rightarrow Cu^{2+} + H_2O_2$$

It is common belief that this enzyme family evolved as a consequence of oxygenation of earth atmosphere, about two billion years ago.

Eukaryotic CuZnSODs are known to be remarkably stable to thermal inactivation and to proteolytic cleavage [1,4]. The eukaryotic enzymes characterized until now, either in the oxidized or in the reduced form, show that not only the molecular fold, but also the subunit association interface, are highly conserved through the eukaryotic phyla [5,6]. Moreover, comparison of the known CuZnSOD amino acid sequences carried out through multiple alignments and based on molecular modeling shows that insertions and deletions of residues occur mainly in the polypeptide loops connecting the β-strands in the barrel, without altering the overall fold [1].

CuZnSOD enzymes have been identified also in prokaryota. Enzymes isolated from the periplasmic space of some bacterial species display a wide and variable pattern of amino acid insertions, deletions and mutations, particularly in the loop regions connecting the expected β-strand segments. Such primary structure variations have been proposed to result in a substantial modification of the subunit interface and in an altered organization of the active site channel, with respect to the eukaryotic CuZnSODs [7]. For instance, the three-dimensional structure of *Photobacterium leiognathi* Cu,Zn SOD reveals a modified quaternary structure assembly for this prokaryotic enzyme, which is however found in the dimeric state. Moreover, spectroscopic studies in solution show that the coordination geometry of the catalytically active copper binding site is different from that observed in the eukaryotic enzymes [8].

Despite their structural differences, all bacterial CuZnSODs have been found to be dimeric [3,9]. However, exceptions to such quaternary structure association are represented by the enzymes from *Escherichia coli* and from *Brucella abortus*, which have been isolated as active and stable monomeric species [3]. Interestingly, comparison of prokaryotic CuZnSODs evolutionary trees highlights the existence of at least two groups of bacterial enzymes, expected to have different monomeric/dimeric aggregation properties [8].

It has been shown that the subunit architecture is conserved independently from the quaternary structure of the enzyme. Monomeric and dimeric CuZnSODs differ in biochemical properties such as catalytic activity, metal affinity, thermal stability, and

susceptibility to protease digestion [10]. Studies on the enzyme from *Salmonella enterica* demonstrated the functional non equivalency of monomeric and dimeric forms. Several strains of this organism possess two independent genes, *sodCI* and *sodCII*, encoding for a dimeric and a monomeric enzyme, respectively [11]. Although the two proteins catalyze the same reaction and share the same subunit fold, the *sodCI* and *sodCII* genes are not functionally interchangeable, indicating that the two proteins have distinctive structural or functional properties, which are apparently related to their different quaternary structure .

To shed more light on the open questions concerning the structural organization of prokaryotic

CuZnSODs, we have characterized via NMR spectroscopy the structure of CuZnSOD from *Salmonella choleraesuis*.

References

[1] D. Bordo, D. Matak, K. Djinovic-Carugo, C. Rosano, A. Pesce, M. Bolognesi, M.E. Stroppolo, M. Falconi, A. Battistoni, A. Desideri, J. Mol. Biol. 285 (1999), 283.

[2] Gabbianelli R., M. D'Orazio, F. Pacello, P. O'Neill, Nicolini L., G. Rotilio, A. Battistoni, Biol. Chem. 385 (2004), 749-754.

[3] D. Bordo, A. Pesce, M. Bolognesi, M.E. Stroppolo, M. Falconi, A. Desideri, John Wiley and Sons, Editor, Handbook of Metalloproteins (Eds.Messerschmidt ed.), Chichester**.** (2001), pp. 1284-1300.

[4] A. Battistoni, M.T. Carri, A.P. Mazzetti, G. Rotilio, Biochem. Biophys. Res. Commun. 186 (1992), 1339-1344.

[5] I. Bertini, C. Luchinat, M. Piccioli, Progr. NMR Spectrosc. 26 (1994), 91-141.

[6] L. Banci, M. Benedetto, I. Bertini, R. Del Conte, M. Piccioli, M.S. Viezzoli, Biochemistry 37 (1998), 11780-11791.

[7] M.E. Stroppolo, A. Pesce, M. D'Orazio, P. O'Neill, D. Bordo, C. Rosano, M. Milani, A. Battistoni, M. Bolognesi, A. Desideri, J. Mol. Biol. 308 (2001), 555-563.

[8] R. Gabbianelli, M. D'Orazio, F. Pacello, P. O'Neill, L. Nicolini, G. Rotilio, A. Battistoni, Biol. Chem. 385 (2004), 749-754.

[9] D. Bordo, K. Djinovic, M. Bolognesi, J. Mol. Biol. 238 (1994), 366-386.

[10] Catacchio B., M. D'Orazio, A. Battistoni, E. Chiancone, In: (2008).

[11] S. Ammendola, M. Ajello, P. Pasquali, J.S. Kroll, P.R. Langford, G. Rotilio, P. Valenti, A. Battistoni, Microbiol 7 (2005), 698-707.

Biomol NMR Assign (2007) 1:65–67
DOI 10.1007/s12104-007-9017-0

ARTICLE

NMR assignment of reduced form of copper, zinc superoxide dismutase from *Salmonella enterica*

Beatriz Jiménez · Mirko Mori · Andrea Battistoni · Marco Sette · Mario Piccioli

Received: 7 February 2007 / Accepted: 6 April 2007 / Published online: 22 May 2007

Abstract Almost complete assignment (97%) of NMR resonances was obtained for the reduced, Cu(I), form of prokaryotic CuZnSOD from *Salmonella enterica*. ^{13}C direct detection was used to complement the standard bouquet of ^{1}H detected triple resonance experiments and contributed to the identification of proline backbone resonances and to side chains assignments of Asx, Glx and aromatic rings. This is the only complete assignment available for monomer SOD from prokaryotic organisms.

Keywords Copper–zinc superoxide dismutase · NMR assignments · ^{13}C direct detection · Metalloproteins · Triple resonance

Biological context

Cu,Zn superoxide dismutases (Cu,ZnSODs) are metalloenzymes which catalyze the disproportion of the toxic superoxide anion into oxygen and hydrogen peroxide. Several Cu,ZnSODs from eukaryotic and prokaryotic sources have been characterized either by X-ray crystallography (Bordo et al. 2001) or NMR spectroscopy (Banci et al. 1998). All these enzymes derive from a common ancestor gene. They display a similar three-dimensional fold, based on a flattened Greek-key eight-stranded β-barrel, and a conserved organization of the active site metal cluster, where a copper ion constitutes the catalytic center and a zinc ion plays mainly a structural role. While eukaryotic Cu,ZnSODs enzymes are homodimers which display a high degree of structural and functional conservation, those isolated from bacterial species show much greater structural variation (Battistoni 2003). Moreover, some prokaryotic Cu,ZnSODs are monomers and it has been shown that variations in the subunit-subunit recognition among different dimeric variants may finely tune Cu,ZnSODs functional properties (Pesce et al. 2000). The only monomeric Cu,ZnSOD characterized so far, by X-ray crystallography, is the *E. coli* enzyme (1ESO.pdb), which displays a catalytic activity comparable to that of dimeric eukaryotic enzymes (Folcarelli et al. 1998).

We report here the resonance assignment of the mature Cu,ZnSOD (amino acids 20-173) from *Salmonella enterica*. This recombinant polypeptide chain has a 46% sequence homology to the dimeric bovine enzyme and 85% sequence identity to the monomeric *E. coli* enzyme. The copper ion was in the reduced, Cu(I), state.

B. Jiménez · M. Mori · M. Piccioli (✉)
Magnetic Resonance Center (CERM) and Department of Chemistry, University of Florence, Via L Sacconi 6, Sesto Fiorentino, Florence 50019, Italy
e-mail: piccioli@cerm.unifi.it

A. Battistoni
Department of Biology, University of Rome ''Tor Vergata'', Rome 00133, Italy

M. Sette
Department of Chemical Science and Technology, University of Rome ''Tor Vergata'', Rome 00133, Italy

Methods and experiments

Recombinant mature CuZnSOD encoded by the *sodC2* gene isolated from a *Salmonella enterica* serovar Choleraesuis strain (Ammendola et al. 2005), was obtained according to a previously reported protocol.

Assignment was performed on uniformly ^{13}C, ^{15}N-labeled derivatives. Samples consisted of about 600 µl solution in standard 5 mm NMR tubes, at 1.5 mM enzyme

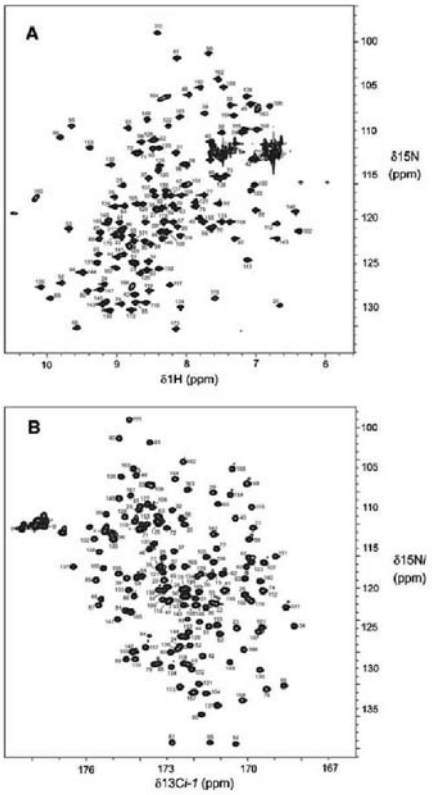

Fig. 1 Panel A: $^{1}H,^{15}N$-HSQC spectrum of monomeric, reduced, Cu(I)ZnSOD from *Salmonella enterica*, recorded at 298 K at 800 MHz. Panel B: $^{13}C,^{15}N$-CON spectrum of the same sample, recorded at 298 K at 176.1 MHz (^{13}C Larmor Frequency). Amino acid labels refer to the assignment of the N resonance. In both panels, side chains assignment is not shown

concentration, in 20 mM phosphate buffer in 95% H_2O/5% D_2O at pH 6.0. Reduced Cu(I) samples were prepared by addition, under oxygen-free atmosphere, of buffered solution of sodium ascorbate in order to obtain a 5 mM concentration of the reductant (Banci et al. 1998).

NMR measurements were performed, at 298 K, on Bruker Avance 800 and Bruker Avance 900 MHz spectrometers, equipped with cryoprobes. Proton chemical shifts are referenced to 2,2-Dimethyl-2-silapentane-5-sulfonate sodium salt (DSS). ^{13}C and ^{15}N chemical shifts are referenced indirectly to DSS using the absolute frequency ratios. Sequence-specific assignment of backbone and side chain resonances was obtained using $^{1}H,^{15}N$-HSQC, $^{1}H,^{13}C$-HSQC, HNCA, HNCO (Kay et al. 1990), HNCACB, CBCA(CO)NH, HBHA(CO)NH (Grzesiek and Bax 1993), (H)CCH-TOCSY, $^{1}H,^{15}N$-NOESY-HSQC. In addition to the carnet of 2D and 3D NMR ^{1}H-detected experiments, we acquired ^{13}C direct detected CACO, CBCACO, CON (Bermel et al. 2006) and CC-COSY (Bertini et al. 2005) experiments. They confirmed Pro sequential assignments and provided complete resonance assignment of Asx and Glx side chains and aromatic rings.

All NMR data were processed using standard Bruker software packages (XWINNMR/TOPSPIN). Spectral analysis and resonance assignment was done using CARA software package (http://www.nmr.ch).

Assignments and data deposition

The amide assignment is summarized in Fig. 1A. Even if ^{13}C direct detection is useful especially for partially or completely unfolded proteins (Bermel et al. 2006), or for systems affected by exchange broadening effects, CON experiment provides an alternative fingerprint also for a well folded protein. Indeed, Figure 1B shows that peak dispersion is comparable, if not better, than a conventional $^{1}H,^{15}N$-HSQC, with the advantage that proline residues are also monitored. A CBCACO experiment allowed us the identification of C^{γ} and C^{δ} carbonyls for Asx and Glx, respectively. Complete sequence-specific assignment of aromatic side chains was obtained using a CC-COSY, which correlates aliphatic and aromatic carbons, combined with a $^{1}H,^{13}C$-HSQC in the aromatic region.

All the atoms observed in routine NMR studies have been identified but Ser and Cys H^{γ}; Thr $H^{\gamma 1}$; Tyr H^{η}; Lys $H^{\varepsilon 2}$, N^{ζ} and H^{ζ}; Phe $H^{\delta 2}$ and $H^{\varepsilon 2}$; Arg C^{ζ} $N^{\eta 1}$, $H^{\eta 11}$ and $H^{\eta 12}$. The overall assignment was 97%. Essentially all ^{13}C resonances were assigned (99.7%), and the score for ^{1}H is also high (96.4 %). The lower percentage of ^{15}N assignments (89.4%) is due to side chain atoms, especially those of Lys N^{ζ}. The assignment of backbone resonances (including also C^{β}) is 100%. The complete resonance list has been deposited on the BMRB (accession number 15112).

Acknowledgments BJ is a Marie Curie Intra European Fellow (EIF) (Contract MEIF-CT-2005-515039). The Consorzio Interuniversitario Risonanze Magnetiche di Metallo Proteine Paramagnetiche (CIRMMP) is gratefully acknowledged for providing access to the high field instrumentation available at CERM, Firenze, under the Project MIUR – FIRB n RBLA032ZM7. We are grateful to Ivano Bertini and Giuseppe Rotilio for their advice and support.

References

Ammendola S, Ajello M, Pasquali P, Kroll JS, Langford PR, Rotilio G, Valenti P, Battistoni A (2005) Differential contribution of *sodC1* and *sodC2* to intracellular survival and pathogenicity of *Salmonella enterica* serovar Choleraesuis. Microbes Infect 7:698–707

Banci L, Benedetto M, Bertini I, Del Conte R, Piccioli M, Viezzoli MS (1998) The solution structure of reduced monomeric Q133M2 Copper, Zinc Superoxide Dismutase. Why SOD is a dimeric enzyme?''. Biochemistry 37:11780–11791

Battistoni A (2003) Role of prokaryotic Cu,Zn superoxide dismutase in pathogenesis. Biochem Soc Trans 31:1326–1329

Bermel W, Bertini I, Felli IC, Piccioli M, Pierattelli R (2006) ^{13}C-detected *protonless* NMR spectroscopy of proteins in solution. Progr NMR Spectrosc 48:25–45

Bertini I, Jiménez B, Piccioli M, Poggi L (2005) Asymmetry in ^{13}C-^{13}C COSY spectra identifies geometry in paramagnetic proteins. J Am Chem Soc 127:12216–12217

Bordo D, Pesce A, Bolognesi M, Stroppolo ME, Falconi M, Desideri A (2001) Cu, Zn Superoxide Dismutase in Prokaryotes and Eukaryotes. In: Messerschmidt A, Huber R, Poulos T, Wieghardt K (eds), Handbook of Metalloproteins. John Wiley and Sons Ltd, Chichester, pp 1284–1300

Folcarelli S, Battistoni A,Falconi M, O'Neill P, Rotilio G, Desideri A (1998) Conserved enzyme-substrate electrostatic attraction in prokaryotic Cu,Zn superoxide dismutases. Biochem Biophys Res Commun 244:908–911

Kay LE, Ikura M, Tschudin R, Bax A (1990) Three-dimensional triple-resonance NMR spectroscopy of isotopically enriched proteins. J Magn Reson 89:496–514

Grzesiek S, Bax A (1993) Amino acid type determination in the sequential assignment procedure of uniformly ^{13}C/15N-enriched proteins. J Biomol NMR 3:185–204

Pesce A, Battistoni A, Stroppolo ME, Polizio F, Nardini M, Kroll JS, Langford PR, O'Neill P, Sette M, Desideri A, Bolognesi M (2000) Functional and crystallographic characterization of Salmonella typhimurium Cu,Zn superoxide dismutase coded by the sodCI virulence gene. J Mol Biol 302:465–478

12954 *Biochemistry* **2008,** *47,* 12954–12963

The Solution Structure of the Monomeric Copper, Zinc Superoxide Dismutase from *Salmonella enterica*: Structural Insights To Understand the Evolution toward the Dimeric Structure[†]

Mirko Mori,[‡] Beatriz Jiménez,[‡,#] Mario Piccioli,[*,‡] Andrea Battistoni,[*,§,||] and Marco Sette[⊥]

Magnetic Resonance Center (CERM) and Department of Chemistry, University of Florence, 50019 Sesto Fiorentino (FI), Italy, Department of Biology, University of Rome "Tor Vergata", 00133 Rome, Italy, National Institute of Biostructures and Biosystems (INBB), Viale delle Medaglie d'Oro 305, 00136 Rome, Italy, and Department of Chemical Science and Technology, University of Rome "Tor Vergata", 00133 Rome, Italy

Received July 3, 2008; Revised Manuscript Received September 19, 2008

ABSTRACT: The structure of the SodCII-encoded monomeric Cu, Zn superoxide dismutase from *Salmonella enterica* has been solved by NMR spectroscopy. This represents the first solution structure of a natural and fully active monomeric superoxide dismutase in solution, providing information useful for the interpretation of the evolutional development of these enzymes. The protein scaffold consists of the characteristic β-barrel common to the whole enzyme family. The general shape of the protein is quite similar to that of *Escherichia coli* Cu, Zn superoxide dismutase, although some differences are observed mainly in the active site. SodCII presents a more rigid conformation with respect to the engineered monomeric mutants of the human Cu, Zn superoxide dismutase, even though significant disorder is still present in the loops shaping the active site. The analysis of both dynamics and hydration properties of the protein in solution highlights the factors required to maintain the fully active and, at the same time, monomeric protein. This study provides novel insights into the functional differences between monomeric and dimeric bacterial Cu, Zn superoxide dismutases, in turn helping to explain the convergent evolution toward a dimeric structure in prokaryotic and eukaryotic enzymes of this class.

Copper, zinc superoxide dismutases (CuZnSOD hereafter)[1] are a class of enzymes that catalyze the disproportion of superoxide radical into oxygen and hydrogen peroxide. Discovered about 40 years ago, they are among the most extensively characterized metalloenzymes in terms of biophysical properties, structure−function relationships, and biological functions (*1*). They have been isolated and studied from a number of eukaryotic and prokaryotic sources (*2*).

Eukaryotic CuZnSODs isolated so far are homodimeric enzymes with an impressive degree of structural homology (*2*). Upon site-directed mutagenesis, it has been possible to obtain monomeric variants (*3*); however, their activity was invariantly lower than that of the native dimeric enzymes (*4*). The availability of mutants of the human enzyme in the monomeric form has facilitated the acquisition of detailed NMR structures (*5*) that, later on, have been extended also to the dimeric form of yeast and human enzymes (*6*). The comparison between solution and solid state structures has led to a better understanding of structural and dynamics processes providing researchers insight concerning structure implication for neurodegenerative diseases such as FALS (*7*).

Prokaryotic CuZnSODs show a much larger structural variability than eukaryotic enzymes, differing either in the active site organization or in the modes of subunit association (*2*). Moreover, some bacterial variants display differences in amino acid sequences, quaternary structures, arrangements of the active site ligands or other species-specific structural/functional properties (*8*−*11*). Interestingly, prokaryotic enzymes can be isolated as active enzymes either as monomers or as dimers with different subunit interfaces (*11*−*14*). The larger structural variability found in prokaryotic enzymes has been addressed via X-ray crystallography (*11, 13, 15*). It has been shown that the subunit architecture is conserved

[†] CIRMMP is gratefully acknowledged for providing access to the high field instrumentation available at CERM, Florence (Contract MIUR-FIRB n.RBLA032ZM7). B.J. is a Marie Curie Intra European Fellow (Grant MEIF-CT-2005-515039).

[*] To whom correspondence should be addressed. Prof. Mario Piccioli, CERM and Department of Chemistry, University of Florence, Via L. Sacconi 3, 50019 Sesto Fiorentino, Italy; tel 39 055 4574265; fax 39 055 4574253; e-mail: piccioli@cerm.unifi.it. Prof. Andrea Battistoni, Department of Biology, University of Rome "Tor Vergata", Rome, Italy; tel 390672594372; fax 39 0672594311.

[‡] University of Florence.

[#] Present address: Laboratory of Structural Biology, Centro de Investigación Príncipe Felipe, Avda. Autopista del Saler 16, 46012 Valencia, Spain.

[§] Department of Biology, University of Rome "Tor Vergata".

[||] National Institute of Biostructures and Biosystems (INBB).

[⊥] Department of Chemical Science and Technology, University of Rome "Tor Vergata".

[1] Abbreviations: CuZnSOD, copper, zinc superoxide dismutase; NMR, nuclear magnetic resonance; FALS, familial amyotrophic lateral sclerosis; SodCII, monomeric Cu, Zn superoxide dismutase from *Salmonella enterica SodCII* gene; SodCI, dimeric Cu, Zn superoxide dismutase from *Salmonella enterica SodCI* gene; Tris, tris(hydroxymethyl)aminomethane; EDTA, ethylenediaminetetraacetic acid; NOESY, nuclear Overhauser effect spectroscopy; HSQC, heteronuclear single quantum coherence; PHOGSY, protein hydration observed by gradient spectroscopy; CLEANEX-PM, phase-modulated CLEAN chemical exchange; AMBER, assisted model building with energy refinement; SDS−PAGE, sodium dodecyl sulfate−polyacrylamide gel electrophoresis.

10.1021/bi801252e CCC: $40.75
Published on Web 11/12/2008

independently of the enzyme quaternary structure. However, monomeric and dimeric CuZnSODs differ in biochemical properties such as catalytic activity, metal affinity, thermal stability, and susceptibility to protease digestion (*16, 17*). Studies on the enzyme from *Salmonella enterica* demonstrated the functional nonequivalency of monomeric and dimeric forms (*17*). Several strains of this organism possess two independent genes, *sodCI* and *sodCII*, encoding for a dimeric (SodCI) and a monomeric (SodCII) enzyme, respectively (*17*). Although the two proteins catalyze the same reaction and share the same subunit fold, the *sodCI* and *sodCII* genes are not functionally interchangeable, indicating that the two proteins have distinctive structural or functional properties, which are apparently related to their different quaternary structure (*16, 17*). The structure of SodCI has been solved by X-ray crystallography (*15*), while SodCII has 85% sequence identity with the monomeric *Escherichia coli* CuZnSOD (*12*) (See Supporting Information, Figure S1).

To the best of our knowledge, no detailed structural characterization for prokaryotic CuZnSODs in solution currently exists. Therefore, in this study, we aimed to solve the solution structure of the mature, monomeric CuZnSOD encoded by the *sodCII* gene isolated from *Salmonella enterica* serovar Choleraesuis strain (SodCII, hereafter). Differences between solid state and solution state can be related to functional aspects and can be a precious source of information to understand reactivity (*5, 18*). Moreover, solution state NMR studies provide unique information on local dynamics that might, in turn, explain functional differences between monomeric and dimeric bacterial and eukaryotic CuZnSODs.

Since the presence of a paramagnetic Cu(II) ion would broaden beyond detection NMR signals within a large sphere around the metal center (*19*), we solved the structure of the reduced Cu(I) state of the protein, which is diamagnetic. We used a so-called hybrid NMR approach, in which the recently developed battery of ^{13}C direct detected experiments (*20*) was used together with routine triple resonance ^{1}H detected experiments to assist NMR assignment and structure determination (*21*). The present structure represents one of the first applications of this tool for structural biologists.

MATERIALS AND METHODS

Protein Expression and Purification. The nucleotide sequence encoding SodCII was inserted in the plasmid pSE420 under control of the *trc* promoter, thus obtaining plasmid pSESCSodCII. This was introduced into *Escherichia coli* BL21 cells to overproduce the protein. Bacteria were cultivated in M9 minimal medium supplemented with 1.5 $\times$ 10^{-5} M $ZnSO_4$ and 4 $\times$ 10^{-6} M $CuSO_4$. ^{15}N-labeled and ^{15}N- and ^{13}C-labeled protein samples were obtained by preparing the minimal M9 medium with ^{15}N ammonium chloride and ^{13}C-glucose purchased from Spectra 2000 (Rome, Italy). Protein overexpression was obtained by inducing the transcription from the *trc* promoter in mid-log-phase cultures by the addition of 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG). After 16 h of growth, cells were harvested by centrifugation and resuspended in 20% sucrose, 30 mM Tris-HCl, pH 8.0, 1 mM EDTA. After incubation on ice in the presence of 1 mg/mL lysozyme, bacteria were centrifuged at 12000*g*, and the supernatant was fractionated by $(NH_4)_2SO_4$ precipitation. Fractions containing CuZnSOD were collected, dialyzed against 150 mM NaCl, 20 mM Tris-HCl, pH 7.4, concentrated to a small volume, and then injected onto a HiLoadTM 16/60 SuperdexTM 75 FPLC gel filtration column (GE-Healthcare). Fractions containing the monomeric protein were concentrated, dialyzed against 20 mM Tris-HCl, pH 7.4, and subjected to ion exchange chromatography on a HiLoad 16/60 FPLC column (Pharmacia) using a 0−0.1 M NaCl linear gradient. Fractions containing CuZnSOD were concentrated and dialyzed against 10 mM phosphate buffer, pH 6.0. Enzyme purity was checked by SDS−PAGE gel electrophoresis.

NMR Experiments. All 2D and 3D spectra were collected at 298 K on Bruker Avance 500, 600, 700, 800, and 900 MHz spectrometers, processed using the TOPSPIN software package (Bruker, Switzerland), and analyzed using CARA software (ETH, Zurich, Switzerland). NMR experiments required to obtain the complete assignment of protein resonances have been previously published (*22*) and are summarized in Supporting Information, Table S2, together with the NMR experiments used to obtain NOE intensities and ^{3}J coupling constants: 2D NOESY, ^{15}N- and ^{13}C-HSQC−NOESY, HNHA, and HNHB.

^{15}N longitudinal and transversal relaxation rates (R_1, R_2), as well as $^{1}H-^{15}N$ NOE values, were measured at 600 MHz, 298 K, as reported in Supporting Information, Table S2. R_1 and R_2 were obtained by fitting peak intensities, measured as a function of the relaxation delay, to a single-exponential decay by using the Levenberg−Marquardt algorithm. Uncertainties had been evaluated by using a Monte Carlo approach. Heteronuclear NOE values were calculated as the ratio of peak intensities in spectra recorded with and without saturation of amide protons. Relaxation data were analyzed using the FASTModelfree software (*23*), which simplifies the input of the Modelfree package.

Hydration properties were studied at 500 MHz via ^{15}N-HSQC−PHOGSY (*24*) and ^{15}N-HSQC−CLEANEX-PM (see Supporting Information, Table S2). All spectrometers, except the 600 MHz spectrometer, were equipped with cryogenically cooled triple resonance ^{1}H probe with a cooled ^{13}C preamplifier (CP-TCI).

Structure Calculation and Refinement. Structure calculation was performed with the software package ATNOS/CANDID/CYANA, using as input the amino acid sequence, the chemical shift lists, and three NMR spectra (two-dimensional NOESY, three-dimensional ^{13}C-resolved NOESY, and three-dimensional ^{15}N-resolved NOESY). The standard protocol was applied using seven cycles of peak picking with ATNOS (*25*), NOE assignment with CANDID (*25*), and structure calculation with CYANA-2.1. Constraints on φ and ψ dihedral angles were derived from the chemical shift index (*26*) and from HNHA, HNHB, and HSQC−NOESY. $^{3}J_{HNH\alpha}$ coupling constants were obtained from the ratio between the intensity of diagonal and cross peaks integrated on $^{1}H-^{1}H$ slices of the HNHA experiment and taking into account the proper correction factor. Scalar coupling constants were converted into dihedral angle constraints by the established Karplus relationship. The ratio between $H_N(i)-H\alpha(i)$ and $H_N(i)-H\alpha(i-1)$ NOEs taken from HSQC−NOESY experiments, provided information on ψ dihedral angles. Overall, 138 φ and 131 ψ dihedral angle values were inserted into the calculation with a tolerance of $\pm 30°$.

Stereospecific assignments were obtained with the program GLOMSA. In the final run, 112 stereospecific assignments were used for calculations. Hydrogen bonds were derived from long-range HNCO experiments (*27*). Overall, 89 hydrogen bonds were used, with H−O distance limits in the 1.8−2.0 Å range. The chemical shift of the Cys55 Cβ atom is indicative of the presence of a disulfide bridge between Cys55 and Cys150. Therefore, the inter-residue Sγ−Sγ distance was supposed to be in the 2.0−2.1 Å range. In each ATNOS/CANDID cycle, angle constraints were combined with the updated NOE upper distance constraints for the subsequent CYANA-2.1 structure calculation cycle. In the final ATNOS−CANDID−CYANA cycle, we obtained 4454 cross-peaks and 3563 assignments from 2D NOESY, 2430 cross-peaks and 2066 assignments from 3D ^{15}N-resolved NOESY, and 3369 cross-peaks and 2751 assignments from ^{13}C-resolved NOESY. At the end of the CANDID module, we had 3291 NOE constraints. After CYANA structure calculation, we obtained a family of 20 conformers with a target function of 10.29 Å and rmsd of 0.63 Å. Using CYANA-2.1, we inserted metal ions Cu(I) and Zn(II) in structure calculations as additional constraints, and 30 conformers were considered for the final calculations. Metal ions within CYANA calculations are defined following an established procedure (*5*). According to the binding topology established from spectroscopic data, copper and zinc ions were linked to the imidazole nitrogen atoms of the coordinated histidine or to the carboxylate oxygen of Asp93, with the N−metal or O−metal distances defined with lower limits of 1.80 Å and upper limits of 2.10 Å (*5*). The 30 conformers with the lowest residual target function values were subjected to restraints minimization with the AMBER 8.0 program (D. Case, UCSF). NOE and torsion angle constraints were applied with force constants of 50 kcal mol^{-1} Å^{-2} and 30 kcal mol^{-1} rad^{-2}, respectively. Force field parameters and charges of the metal ions and their ligands were the same as those used for restrained energy minimization and dynamics of reduced form of human dimeric protein (*28*). Structure quality was evaluated using the program PROCHECK-NMR and WHATIF. Structure visualization was done with the program MOLMOL (ETH, Zurich, Switzerland). Structure calculations and analysis were performed on a cluster of Linux PCs.

Atomic coordinates were deposited to the RCSB Protein Data Bank, accession number 2k4w.

RESULTS

NMR Spectral Assignment. Throughout this paper, we have used the amino acid numbering of the mature protein in which the first 19 residues have been deleted. Sequence-specific assignment of backbone and side chain resonances (*22*) was obtained using HSQC, HNCA, HNCO, HNCACB, CBCA(CO)NH, HBHA(CO)NH, (H)CCH-TOCSY, and ^{1}H,^{15}N-NOESY−HSQC. In addition to the set of 2D and 3D ^{1}H-detected experiments, we acquired ^{13}C direct detected CACO, CBCACO, CON and CC-COSY experiments (*20, 29*−*31*). They confirmed proline sequential assignments and provided complete resonance assignment of aspartic acid/asparagine and glutamic acid/glutamine side chains and aromatic rings. Even if ^{13}C direct detection is especially useful for partially or completely unfolded proteins (*32*) or for systems affected by exchange broadening effects, CON experiments provide an alternative fingerprint also for a well-folded protein (*22*).

All atoms observed in routine NMR studies have been identified except serine and cysteine Hγ, threonine Hγ1, tyrosine Hγ, lysine Hε2, Nη, and Hη, phenylalanine Hδ2 and Hε2, and arginine Cη, Nγ1, Hγ11, and Hγ12. The overall assignment was 97%. Almost all ^{13}C resonances were assigned (99.7%), and also the score for ^{1}H was high (96.4%). The lower percentage of ^{15}N assignments (89.4%) was due to side chain atoms, especially those of lysine Nη. The assignment of backbone resonances, including Cβ, was 100%. The complete resonance list has been deposited on the BMRB (accession number 15112) (*22*).

The topology of metal bound histidines has been unambiguously determined from the ^{1}H NMR spectrum of the exchangeable histidine imidazole protons and HSQC patterns (Supporting Information, Figure S3) and from NOESY experiments. NOESY cross peaks from histidine side chains to backbone atoms provide the residue specific assignment of side chain imidazole groups and information on the metal binding topology (*3, 33*). His50 and His128 are bound to the metal via Nε2, while His48, His73, His82, and His90 are bound to the metal via Nδ1. His36 and His45 do not bind any metal ion. The comparison with existing CuZnSOD structures suggests that His48, His50, and His128 are bound to Cu^{+}, while His73, His82, and His90 are bound to Zn^{2+}. Even if there is no direct spectroscopic evidence for the existence of an aspartic acid COO^{-}−Zn^{2+} bond, the presence of a conserved aspartic acid at position 93 that is invariantly bound to Zn^{2+} suggests that Asp93 COO^{-} is Zn-bound, as in the active site of all CuZnSODs characterized thus far. Furthermore, the occurrence of a direct COO^{-}−Zn^{2+} bond is supported by a number of inter-residue NOEs experienced by Asp93 Hβ protons. These NOEs indicate that Asp93 side chain is located in the active site, pointing toward the Zn^{2+} ion.

General Description of the Structure. After restrained energy minimization with the AMBER 8.0 program, the final bundle of 30 conformers of SodCII has an average total target function of 0.60 ± 0.08 Å^2 (CYANA units). The rmsd (residues 3−150) from the average structure is 0.96 ± 0.12 Å for the backbone atoms and 1.30 ± 0.12 Å for the heavy atoms. In Table 1, the statistics on constraint violations in the final family and in the average structure are reported, together with selected quality parameters from a PROCHECK-NMR and WHATIF analysis. These data indicate that the solution structure obtained for SodCII is of good quality. According to the Ramachandran plot (Table 1), 95% (3305 residues) of the non-glycine and non-proline residues have φ dihedral angle in allowed regions, 3.0% (105 residues) in the generously allowed regions, and 2% in disallowed regions (70 residues).

The family of structures obtained for SodCII is shown in Figure 1A (backbone and metal ions only). The structure shows the typical CuZnSOD eight-stranded β-barrel Greek key, conserved in all the conformers (β-strands are depicted in violet in Figure 1A, while the seven loops are in pink). Two small antiparallel β-strands for residues 59−60 and 63−64 in loop IV (cyan) and two short α-helices for residues 117−120 and 141−143 (red) are also observed within loop VI and loop VII, respectively. The α-helix from residues 117−120 is also conserved throughout the family, while the

Table 1: Statistical Analysis of the Energy-Minimized Family of Conformers and Mean Structure of SodCII

	Sod-CII (30 structures)	Sod-CII (mean)
rms Violations per Meaningful Distance Constraints (Å)[a]		
intraresidue (339)	0.0104 ± 0.0014	0.0109
sequential (947)	0.0047 ± 0.0015	0.0040
medium range (391)	0.0026 ± 0.0026	0.0000
long range (1285)	0.0048 ± 0.0011	0.0033
Total (3022)	0.0056 ± 0.0008	0.0048
rms Violations per Meaningful Angle Constraints (deg)		
φ (210)	1.40 ± 0.97	2.1957
ψ (213)	1.59 ± 1.15	0.8915
avg no. constraints per residue	21	
Average Number of Violations per Conformer		
φ	1.87 ± 1.17	2
ψ	1.67 ± 0.94	1
NOE violations larger than 0.3 Å	0.000 ± 0.00	0
NOE violations between 0.1 and 0.3 Å	1.5 ± 1.2845	8
Average rmsd to the Mean (Å)[b]		
backbone	0.96 ± 0.12 Å	1.30 ± 0.12 Å
heavy		
Structural Analysis[c]		
% of residues in most favorable regions	70.5	72.4
% of residues in allowed regions	24.5	22.4
% of residues in generously allowed regions	3	2.6
% of residues in disallowed regions	2	2.6
WHAT IF Structure *Z*-Scores[d]		
1st generation packing quality	−3.195	−3.135
2nd generation packing quality	−1.752	−0.745
Ramachandran plot appearance	−3.617	−2.739
χ_1/χ_2 rotamer normality	−2.878	−2.459
backbone conformation	−3.784	−4.848
WHAT IF rms *Z*-Scores[e]		
bond lengths	0.662	0.653
bond angles	1.303	1.253
omega angle restraints	2.011	2.089
side chain planarity	2.004	1.813
improper dihedral distribution	1.068	1.003
inside/outside distribution	0.928	0.913

[a] Calculated over the 30 conformers representing the NMR structure. The mean value and the standard deviation are given. [b] Calculated for the backbone atoms N, C^{α}, and C′ of residues 3−150. [c] Ramachandran plot analysis performed with PROCHECK. [d] Calculated on all residues with the exception of those located in loop IV. A *Z*-score is defined as the deviation from the average value for this indicator observed in a database of high-resolution crystal structures, expressed in units of the standard deviation of this database-derived average. Typically, *Z*-scores below a value of −3 are considered poor, those below −4 are considered bad [Vriend, G. (1990) J. Mol. Graphics 8]. [e] Values calculated on all residues with the exception of those located in loop IV.

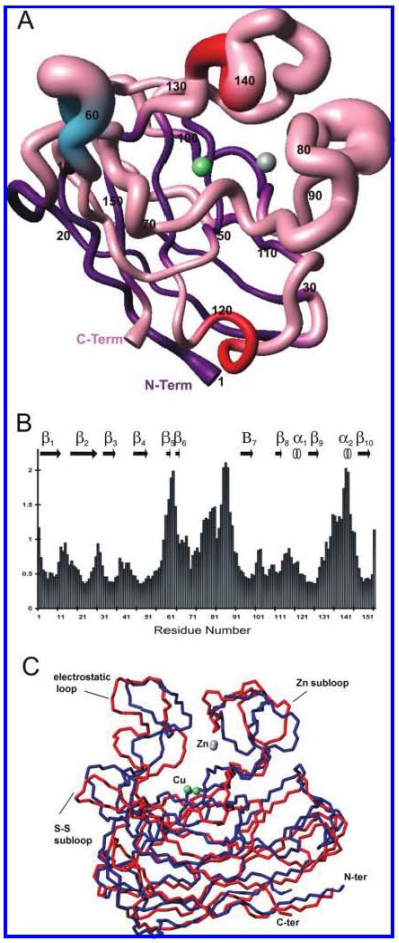

FIGURE 1: (A) Family of 30 accepted structures of reduced SodCII, displayed with MOLMOL. Structures were aligned according to best-fit superimposition of C, Cα, and N atoms of residues 3−150. The metal ions are pictured in green (copper) and in gray (zinc), the β-barrel in purple, the two small β-strands in blue and the two α-helices in red. (B) Backbone rmsd values for the best 30 conformers of SodCII. (C) Superposition of the Cα trace of the mean structure of reduced SodCII (red) with X-ray structure of *E. coli* SOD (blue) (1eso.pdb).

141−143 α-helix is conserved only for 23 out of 30 conformers.

The rmsd for the β-barrel (Figure 1B) is 0.56 ± 0.08 and 1.08 ± 0.0.15 Å for backbone and all heavy atoms, respectively. The most disordered regions, where backbone rmsd exceeds 1 Å, extend from Ala58 to Ala64, from Asp75 to Gly89, and from Gly131 to Gly145. Region 58−64 also includes the two β-strands that do not belong to the β-barrel. Region 75−89, which has the largest rmsd (ca. 2.5 Å), encompasses the Zn bound ligands from His73 to His90. Both regions 58−64 and 75−89 belong to the long loop IV, which, together with loop VI and loop VII, provides the architecture to build the metal binding site and the active cavity. Also the Gly131−Gly145 region is critical for CuZnSODs, because this segment participates in the construction of the active site.

A comparison of the NMR structure of SodCII with the X-ray structure of the *E. coli* enzyme, depicted in Figure 1C, identifies only minor differences in the β-barrel. The topology of loop IV, the longest loop of the enzyme, is well conserved in the two structures. Indeed, the disulfide bridge

between Cys55, at the beginning of loop IV, and Cys150 in strand 8 contribute to anchor loop IV with respect to the β-barrel, therefore favoring the typical loop topology that gives rise to the zinc binding pocket. The short α-helix identified in *E. coli* CuZnSOD between residues 64 and 67 is not apparent in SodCII, where it is instead replaced by two short β-strand segments.

Active Site Channel. Residues that define the access to the active site channel are located in the first part of loop IV, also termed the S−S subloop (residues 51−73), and in loop VII, referred to as the electrostatic loop (residues 129−146) in the eukaryotic enzymes. For all of them, the local backbone rmsd is larger than the average value. As observed in other bacterial CuZnSODs (*14*), these loops display major structural differences with respect to eukaryotic enzymes. In bacteria, the S−S subloop region has an eight-residue insertion, whereas loop VII shows a four-residue deletion. The longer S−S subloop folds at the entrance of the active site channel and contributes to reduce solvent exposure. In SodCII, this loop conserves the charged residue motif (KDGK) typical of most bacterial CuZnSODs (*2*). These residues are located in front of the active site channel and are involved in the electrostatic substrate attraction toward the active site, thus functionally substituting the charged residues located in loop VII in the eukaryotic enzymes (*34*). Unlike the dimeric bacterial enzymes, the S−S loop of SodCII and *E. coli* CuZnSOD contains an additional glutamic acid residue at position 68.

A conserved residue within the electrostatic loop is Arg147 (*2*). This is a key residue for the activity because it is either involved in the electrostatic attraction mechanism or in a direct interaction with the substrate (*3, 5*). The role and the orientation of its guanidinium group has been proposed to be a driving factor for the enzymatic activity (*5*). In human oxidized CuZnSOD (*35*), three H-bonds with neighboring residues locate the side chain of Arg147, while in *Xenopus laevis* CuZnSOD (*36*) the correct orientation is mainly given by hydrophobic interactions between its side chain and neighboring residues. In *E. coli* CuZnSOD, the orientation of Arg147 is imposed by a salt link between its guanidinium group and Glu68 (*37*), and it is thought that this interaction can compensate the absence of other stabilizing contacts that occur in eukaryotic CuZnSODs (*12*). In the present structure, Arg147 is not involved in any H-bond network similar to that characterizing human CuZnSOD (*38*). We observe three residues, Ala66, Met126, and Val10, whose side chains are close to the aliphatic part of Arg147 with inter-residue H−H distances between 3.80 and 4.75 Å. This suggests that hydrophobic interactions among the aliphatic side chain of Arg147 and nearby residues play a major role for the optimal orientation of the arginine residue as previously observed in *X. laevis* CuZnSOD (*2*). From the ensemble of structures, we have found two different conformations for Arg147. One of them is closer to the active site and corresponds to the position present in the average structure, whereas the other one is more solvent-exposed. The average structure suggests an interaction between Arg147 guanidinium groups and Gly145 and Ala67 carbonyl oxygens (147 Nη1−145 O = 2.83 Å; 147 Nη2−67 O = 2.78 Å). A very similar orientation is found in *E. coli* CuZnSOD.

Metal Binding Residues. The average rmsd calculated over metal binding residues, considering all heavy atoms, is

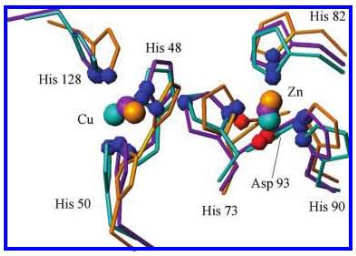

FIGURE 2: The active site superimposition of SodCII (orange), *E coli* CuZnSOD (purple), and human CuZnSOD (green-blue).

smaller than 1 Å, and it is lower than the average global value observed for side chains. Indeed, metal-to-donor coordination bonds and several inter-residue NOEs among histidines contribute to the precision of the active site structure. The superimposition of backbone atoms of the metal ion ligands in SodCII and *E. coli* CuZnSOD, shown in Figure 2, exhibits an rmsd value of 0.47 Å.

Zn-bound His73 is protonated at the Nε2 position, consistent with a broken bridge and an imidazole ring tilted with respect to the oxidized form. The distance between copper and HNε2 is 3.15 Å, and it is larger than the van der Waals distance, while the distance between the copper ion and His73Nε2 is 3.28 Å. A similar Cu−N distance (3.44 Å) has been observed in the human reduced enzyme solved by NMR (*5*). This is also consistent with the X-ray structure of the fully reduced yeast enzyme, in which the Cu-HisNε2 is 3.2 Å, longer than that observed in oxidized CuZnSODs (2.1 Å in *X. laevis*, for example) (*39*). In contrast, intermediate distances were found in the crystal structure of oxidized bacterial enzymes from *E. coli* (2.7 Å) (*12*) and SodCI (2.5 Å) (*15*). As previously suggested (*40*), in the prokaryotic enzymes solved by X-ray crystallography, a partial but not complete reduction of the copper ion occurred under the X-ray beam.

The oxidation state of the copper ion is possibly responsible for some of the differences observed between SodCII and *E. coli* SOD. Among them, the His73 ring plane flips by approximately 20° with respect to the position occurring in *E. coli* when the bridge between the two metal ions is present. This movement is similar to that previously revealed by NMR in the case of the human enzyme and is required to prevent the occurrence of van der Waals violations when His73 is protonated at the Nε2 position.

A slight shift is also observed for His128, one of the copper ion ligands. In SodCII, the His128 ring plane moves toward His48 with respect to the position observed in the *E. coli* enzyme. As a consequence, the distance between His128 Cε1 and His48 Nδ1 is 3.79 Å, while in *E. coli* CuZnSOD it is 4.29 Å. However, the Cu−Zn distance is 6.6 Å, which is similar to that observed in *E. coli* CuZnSOD (6.5 Å) and in most of the available X-ray and NMR structures of other CuZnSODs (*2*). Even though NMR data do not provide direct constraints on the Cu−Zn distances, and the latter is driven not only by the experimental NMR restraints but also from the additional links imposed from the histidine donor atoms to the metal, these data indicate that the metal−metal distances are not affected by metal ion reduction. Another

Table 2: Interface Residues in PlSod, ApSod, SodCI, SodCII, and *E.coli* Sod[a]

	2,3 loop		β6d	Zn subloop	β5e	β4f		7,8 loop	
	27−29	32	44	84, 85B, 87	96−98	107−110	112, 113	132	134
PlSod	NK**Y**	V	M	F**W**D	ALF	**N**PVL	**PR**	**D**	H
ApSod	SA**Y**	V	L	Y**W**D	ALF	**N**PVL	**PR**	**D**	H
SodCI	TP**Y**	L	I	GY**D**	GLV	YPLL	**PR**	**D**	Y
SodCII	TEG	K	E	G − G	VLV	EPVT	PR	D	M
E. coli Sod	TDK	E	E	G − G	ALV	DAVI	PR	D	M

[a] The fully conserved residues in dimeric Sod are in bold. Numbering refers to the SodCII studied in this work. Both SodCII and *E. coli* Sod have a deletion at position 85B.

crucial residue to build the metal binding site is Asp132. This residue drives the orientation of Cu-bound His48 and Zn-bound His82 forming distinct H-bonds with imidazole HNε2 protons, as occurs in other CuZnSODs.

Comparison between SodCII and Dimeric Bacterial CuZn-SODs. The overall molecular fold of SodCII is comparable to that of the subunits of the dimeric enzymes from *Salmonella enterica* (SodCI), *Photobacterium leiognathi* (PlSod) (*39*), and *Actinobacillus pleuropneumoniae* (*40*), which display 62%, 55%, and 62% sequence identity, respectively, to SodCII. In these enzymes, the association of the two subunits is mostly promoted by van der Waals interactions and by a minor contribution from hydrogen bonds (*20, 38*).

In particular, five regions are important for dimeric assembly and stabilization of quaternary structure: loop II, Zn subloop, β-strands 4 and 5, and loop VII. These protein fragments are characterized by some evolutionarily invariant residues that are involved in subunit−subunit interaction. Table 2, where residues are numbered according to SodCII sequence, shows residues at the interface that give rise to three clusters of interactions that drive the dimer stabilization. There are substantial differences between SodCI and SodCII. Four tyrosine residues, Tyr29, Tyr85B, Tyr107, and Tyr134, are missing or substituted with non-aromatic residues in SodCII. The Tyr29→Gly mutation, as well as the Tyr85B deletion, severely impair many of the van der Walls hydrophobic contacts between the two subunits of SodCI. In SodCI, the cluster region is disrupted by residues 84 and 85B on one side, and residues 27, 29, 32, and 108−110 on the other. Additionally, the Tyr29→Gly replacement, together with the Asp87→Gly substitution, prevents the formation of a Tyr29−Asp87 H-bond, which has been described as a crucial factor for the subunit association. Likewise, Tyr107→Glu, together with Ile44→Glu substitution, provides a negatively charged region that prevents van der Walls contacts in the second cluster region, defined by residues 44, 98, and 107. Finally, the hydrophobic cavity that normally hosts the buried water molecule bound to three tyrosine residues, is clearly destabilized by the substitution of the Tyr134 by methionine.

A similar situation applies to the other monomeric protein, that is the *E. coli* enzyme, where the same substitutions described above (Tyr29, Tyr85B, Tyr107, Tyr134, Ile44, and Asp87) are observed.

Thus, the replacement of a few hydrophobic residues and the increase of residues with different size and charge throughout the hypothetical intersubunit region give rise to a surface that is nonconducive to the formation of homo- or heterodimeric proteins.

Protein Hydration. The analysis of ePHOGSY and CLEAN-EX-PM experiments (*24*) shows the occurrence of water molecules interacting with specific protein regions at different exchange rates. Strong CLEANEX-PM peaks are observed for a few amino acids, namely, residues 3, 12, 18, 28, 31, 49, 61, 63, 67, 72, 101, 145, and 146. They point out the protein regions in which H_2O molecules bind to the protein in an exchange regime faster than the rotational correlation time of the protein. As shown in Figure 3, most of these amino acids (8 out of 13) belong to the active channel. They provide the hydrophilic environment contributing to driving superoxide toward the metal ion. The fast exchange rate experienced by these H_2O molecules is required in order to allow the fast diffusion of substrate to the active site.

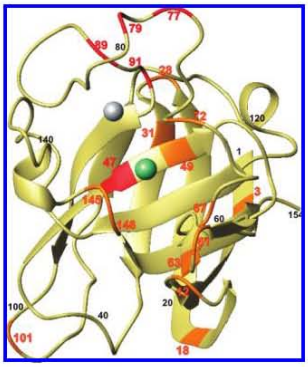

FIGURE 3: SodCII cartoon in which the residues identified via CLEANEX-PM and PHOGSY experiments have been highlighted. Residues colored in yellow (aa 3, 12, 18, 31, 49, 61, 63, 67, 72, 101, 145, and 146) are observed in the CLEANEX-PM experiments. The exchange rate of amide protons with water molecules is faster than the rotational correlation time of the protein. Residues colored in red (aa 47, 77, 79, 89, and 91) are identified via the ePHOGSY experiment. The water−protein interaction is therefore slower than the rotational correlation time (structural water molecule).

The ePHOGSY peak to His48 HNε2 accounts for the presence of a buried water molecule close to the copper ion (*41*). However, the HOH−Cu distance, which is known to be different in various CuZnSODs (*2*), cannot be accurately estimated from this measurement. Another buried water molecule is observed within the Zn-binding loop, indicated by ePHOGSY peaks involving amino acids 77, 79, 89, and 91. However, no ePHOGSY peaks are observed from the Zn-coordinated imidazole NH groups. This suggests that the H_2O molecule observed in proximity of the Zn binding loop is not inside the active cavity. Indeed, residues 75−80 and 85−92 define a loop where a water molecule could be buried due to the presence of H-bonds with Asn78 and Gly89

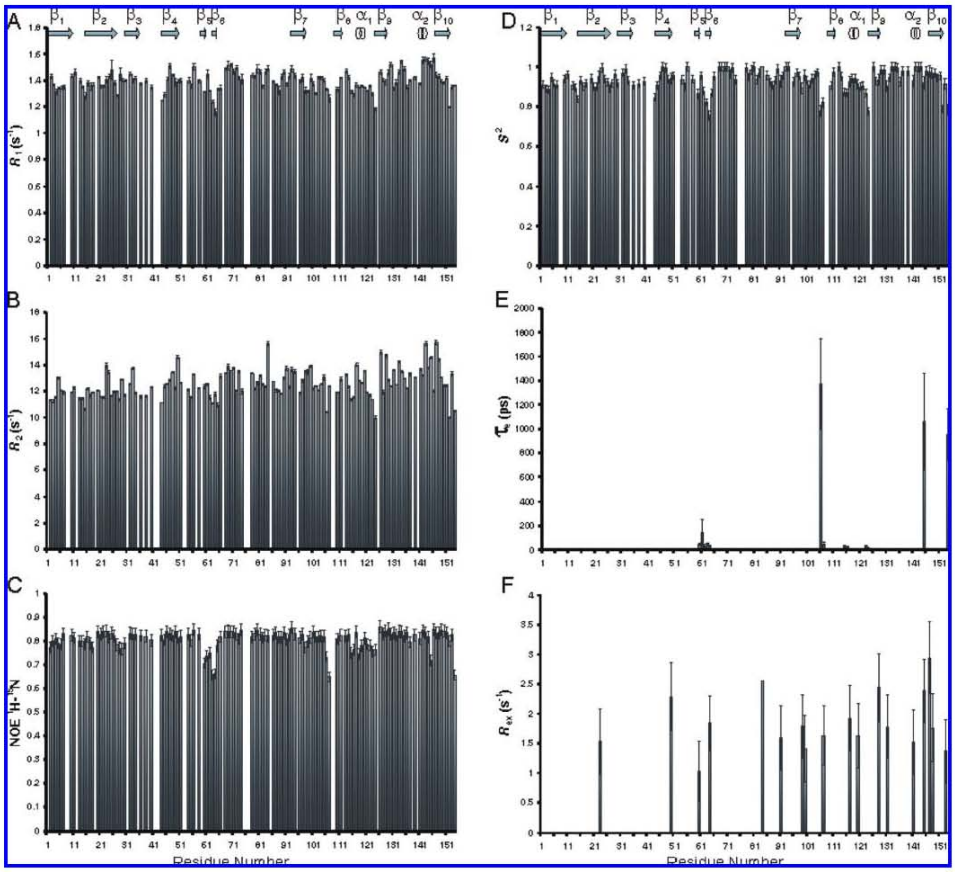

FIGURE 4: Plots showing the relaxation rates and mobility parameters in reduced SodCII, meassured at 600 MHz and 298 K: (A) longitudinal relaxation; (B) transverse relaxation; (C) NH NOE; (D) order parameter, S^2; (E) local correlation time for motions on the subnanosecond time scale (τ_s); (F) exchange rates on the millisecond−microsecond time scale (R_{ex}).

contributing to the stabilization of the secondary structure of the Zn-binding loop.

Protein Dynamics in Solution. Relaxation properties of 126 out of 142 assigned $^1H-^{15}N$ resonances were analyzed. Average values of 1.4 ± 0.1 and 12 ± 1 s^{-1} were obtained for ^{15}N longitudinal (R_1) and transverse (R_2) relaxation rates, respectively, while the average $^1H-^{15}N$ NOE value was 0.80 ± 0.04 s^{-1}. A simple isotropic model was sufficient to fit the data using a τ_c of 8.99 ± 0.02 ns. This value is similar to that observed for the monomeric variant of the human superoxide dismutase (*18*), as well as other proteins of its size. The average order parameter value ($S^2 = 0.94 \pm 0.04$) confirms that the protein fold maintains a rigid structure. This value is closest to the dimeric form of the homologous human form ($S^2 = 0.92 \pm 0.09$) compared with that of the monomeric one ($S^2 = 0.89 \pm 0.11$) (*18*), the latter being significantly less active (*5*). Relaxation data for SodCII are summarized in Figure 4. The regions encompassing the two shortest β-sheets in the protein (Ile59−Ala64), the fragment involving Thr106−Glu107, and Ala124 display a faster internal mobility. The two β-sheets in the 58−64 region form a subloop that is quite solvent-exposed (as pointed out by hydration experiments, in which peaks are observed for aa 61 and 63) and that defines the more peripheral part of the active channel. However, low H−N NOEs and small R_2 values suggest that the intervening motions are on the order of subnanoseconds and that no large conformational exchange motions (typically at milli- or microsecond time scale) occur.

DISCUSSION

All the structures of prokaryotic SOD described to date have been obtained via X-ray crystallography. In this frame, the solution structure of SodCII allows us to gain insights into the dynamic and hydration properties of the protein and

to correlate such properties with the enzymatic activity. Recently, it has been observed that the catalytic activity of SodCI is significantly higher than that of SodCII (*17*) and that the activity of SodCII is comparable to that of the monomeric enzyme from *E. coli* (Battistoni et al., manuscript in preparation). The catalytic rate of monomeric *E. coli* and SodCII enzymes, measured by pulse radiolysis at low ionic strength, is about 1.3×10^9 M^{-1} s^{-1} (*34*). This value is comparable to that of the dimeric human enzyme (2.0×10^9 M^{-1} s^{-1}) but is considerably lower than those of the dimeric SodCI (1.3×10^{10} M^{-1} s^{-1}) or PISod (8.5×10^9 M^{-1} s^{-1}) bacterial enzymes (*16*). Moreover, the activity of the natural monomeric prokaryotic enzymes is much higher than those reported for the engineered monomeric human enzymes (about 2.0×10^8 M^{-1} s^{-1}) (*42*) whose large decrease in catalytic rate is a likely consequence of the alterations in the tertiary structure caused by extensive solvation or distortion of the mutated dimer interface.

Our relaxation data allow speculation concerning the activity of the enzyme and the monomer versus the native dimer state. When relaxation data of dimeric vs monomeric mutants of human SOD are compared, motions in the millisecond time scale have been proposed to be responsible of multiple conformations, which eventually prevent the active site residues from assuming a functional orientation (*18*). As no evidence for enhanced local mobility nor conformational exchange on the millisecond time scale is present in SodCII, in contrast to that observed for the partially active monomeric human enzyme, the dynamics of SodCII on the slow time scale resembles the fully active eukaryotic dimeric enzyme rather than the partially active monomeric enzyme. This supports the notion that fluctuations of the electrostatic loop may affect the diffusion of substrate toward the active site and therefore decrease the activity.

The mobility observed on fast time scales for the S−S subloop (see Figure 4) is in excellent agreement with a previous study on the *E. coli* enzyme based on the combined use of limited proteolysis and molecular dynamics, showing that this loop is quite mobile on the nanosecond−picosecond time scale (*43*). The solution structure of dimeric human CuZnSOD reveals an increase in mobility of the loops shaping the active site thus pointing out that subnanosecond mobility facilitates the diffusion of the substrate through the active channel. The solution structure of human CuZnSOD has also revealed that metal ions play a role in protein structure (*44, 45*). Moreover, several studies have suggested that in monomeric bacterial CuZnSODs the mobility of the S−S loop might be affected more than in dimeric ones by metal occupancy of the active site (*17, 46*).

The stability of monomer in solution is favored by the high average S^2 observed. The rigidity of SodCII, together with unfavorable electric charge distribution and the substitution of hydrophobic residues throughout the hypothetical intersubunit region, contributes to the stability of the monomer in solution. It has been already shown that enhanced mobility of those amino acids involved in the subunit−subunit interactions is one of the crucial factors for the aggregation of dimeric CuZnSODs (*7*).

Another relevant result from this work is the identification of water molecules in the active site channel. Thus, the occurrence of increased mobility of the loops shaping the active site and the presence of charged residues important for electrostatic attraction cooperate with the presence of water molecules that can help the diffusion of the substrate in order to obtain high catalytic efficiency.

These observations are relevant to understand the structural and functional differences between monomeric and dimeric enzymes. All characterized eukaryotic CuZnSODs show a tight and stable dimeric structure. The amino acids involved in dimer formation are highly conserved and the subunit−subunit recognition is based both on hydrophobic interactions and direct hydrogen bonding (*5, 13*). The highly conserved dimeric structure, typical of eukaryotic SODs, is not observed in prokaryotic enzymes. Indeed, some enzymes of this class, such as SodCII, are monomeric, while others are dimeric yet present different subunit arrangements (*15*). The residues implicated in the interface subunit in the dimeric prokaryotic enzymes are not always conserved. The surface is much less hydrophobic when compared with the eukaryotic enzymes, and it is characterized by the presence of intervening water molecules, which can be considered components of the bacterial quaternary structure, located within an intersubunit cavity (*2, 14, 15, 47*). Mutations in the interface residues and variations in number and orientation of trapped water molecules may support a fine-tuning of subunit association processes across the different prokaryotic species. Single amino acid mutations at the dimer interface of PISod significantly decrease the affinity for copper but increase catalytic rates, possibly due to an enhanced copper accessibility and to an altered mobility of the S−S loop (*15, 43*). Species-specific modifications in the intersubunit region have therefore been proposed to contribute also to the exceptionally high activity of SodCI (*15*) and to the differences in dimer stability, metal affinity, and protease resistance between SodCI and PISod (*16, 17*).

The monomeric enzymes from *E. coli* and *S. enterica* have clearly distinct functional features when compared with the dimeric bacterial enzymes. These include lower activity and thermal stability, looser metal cofactor affinity, and decreased resistance to protease digestion (*16, 17*). These differences have been tentatively explained as a consequence of alterations in the mobility of the loops forming the active site channel. Similar properties have been observed in PISod interface mutants, which have a strong tendency for monomerization (*2*). The present investigation shows that monomeric SodCII is characterized by significant disorder in the loops shaping the active site, thus supporting this hypothesis.

The dimer interface of prokaryotic and eukaryotic CuZnSODs involves different β-strand elements, indicating an independent evolution toward the dimeric structure (*14*). Comparative studies on naturally occurring monomeric and dimeric bacterial enzymes show that the dimeric structure is associated with higher stability and activity (*16, 17*). This is highlighted by recent studies on the two *Salmonella* enzymes, which have shown that *sodCI* and *sodCII* are not functionally exchangeable (*17*). In fact, when the sequence encoding for SodCII is placed under the control of the *sodCI* promoter, it is not able to complement the lack of *sodCI*, therefore suggesting that the dimeric enzyme has superior properties. This observation raises questions regarding the meaning of monomeric enzymes in different bacteria. Monomers cannot be trivially considered as evolutionary relics still present in the genome of scattered bacteria. In fact, bacteria such as *Salmonella* and *E. coli O157:H7*

express either monomeric or dimeric variants of the enzyme (*48*). Moreover, dimeric CuZnSODs can be isolated from the very ancient hyperthermophilic eubacterium *Aquifex aeolicus* (Battistoni, A., and D'Orazio, M., unpublished observations), indicating that dimeric structure formation in CuZnSOD was an early step during evolution. In fact, SodCI has a clear role in bacterial protection from phagocytic attack (*17*), while the exact role of SodCII is still poorly defined. Although it has been proposed that this enzyme protects bacteria from toxic effects of endogenous superoxide generated within the periplasmic space (*49*), additional studies are required to clarify whether its functions are correlated with the monomeric structure.

In summary, this study describes the first solution structure of a natural monomeric CuZnSOD from *Salmonella enterica*. We provide indications about the mobility and the hydration of the enzyme and clarify the essential factors required to obtain a stable and fully active monomeric enzyme. In addition, the structural and mobility features of the protein surface highlight differences between monomeric and dimeric CuZnSODs, which help to explain the functional diversification between the bacterial enzymes of this class and the convergent evolution of bacterial and eukaryotic CuZnSODs toward the dimeric structure.

ACKNOWLEDGMENT

We thank Prof. I. Bertini and Prof. G. Rotilio for critical reading and continuous support. We acknowledge Dr. F. Cantini for sharing her expertise in CARA and CYANA software. We are grateful to Dr. D. A. MacIntyre for careful readings of this manuscript.

SUPPORTING INFORMATION AVAILABLE

Multiple alignment of amino acid sequence of CuZnSODs, HSQC spectrum of imidazole groups, and acquisition parameters for NMR experiments. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Valentine, J. S., Doucette, P. A., and Potter, S. Z. (2005) Copper-Zinc Superoxide Dismutase and Amyotrophic Lateral Sclerosis. *Annu. Rev. Biochem. 74*, 563–593.
2. Bordo, D., Pesce, A., Bolognesi, M., Stroppolo, M. E., Falconi, M., and Desideri, A. (2001) *Handbook of Metalloproteins*, pp1284−1300, John Wiley and Sons, Chichester.
3. Bertini, I., Piccioli, M., Viezzoli, M. S., Chiu, C. Y., and Mullenbach, G. T. (1994) A spectroscopic characterization of a monomeric analog of copper-zinc superoxide dismutase. *Eur. J. Biophys. 23*, 167–176.
4. O'Neill, P., Davies, S., Fielden, E. M., Calabrese, L., Capo, C., Marmocchi, F., Natoli, G., and Rotilio, G. (1988) The effects of pH and various salt upon the activity of a series of superoxide dismutases. *Biochem. J. 251*, 41–46.
5. Banci, L., Benedetto, M., Bertini, I., Del Conte, R., Piccioli, M., and Viezzoli, M. S. (1998) Solution structure of reduced monomeric Q133M2 copper, zinc superoxide dismutase. Why is SOD a dimeric enzyme? *Biochemistry 37*, 11780–11791.
6. Banci, L., Bertini, I., Cantini, F., D'Amelio, N., and Gaggelli, E. (2006) Human SOD1 before harboring the catalytic metal: Solution structure of copper-depleted, disulfide-reduced form. *J. Biol. Chem. 281*, 2333–2337.
7. Banci, L., Bertini, I., Girotto, S., Martinelli, M., Vieru, M., Whitelegge, J., Durazo, A., and Valentine, J. S. (2007) Metal-free SOD1 forms amyloid-like oligomers: A possible general mechanism for familial ALS. *Proc. Natl. Acad. Sci. U.S.A. 104*, 11263–11267.
8. Pacello, F., Langford, P. R., Kroll, J. S., Indiani, C., Smulevich, G., Desideri, A., Rotilio, G., and Battistoni, A. (2001) A novel heme protein, the Cu,Zn-superoxide dismutase from *Haemophilus ducreyi*. *J. Biol. Chem. 276*, 30326–30334.
9. Banci, L., Bertini, I., Calderone, V., Cramaro, F., Del Conte, R., Fantoni, A., Mangani, S., Quattrone, A., and Viezzoli, M. S. (2005) A prokaryotic superoxide dismutase paralog lacking two Cu ligands: From largely unstructured in solution to ordered in the crystal. *Proc. Natl. Acad. Sci. U.S.A. 102*, 7541–7546.
10. Battistoni, A., Pacello, F., Mazzetti, A. P., Capo, C., Kroll, J. S., Langford, P. R., Sansone, A., Donnarumma, G., Valenti, P., and Rotilio, G. (2001) A histidine-rich metal binding domain at the N terminus of Cu,Zn-superoxide dismutases from pathogenic bacteria. *J. Biol. Chem. 276*, 30315–30325.
11. Spagnolo, L., Toro, I., D'Orazio, M., O'Neill, P., Pedersen, J. Z., Carugo, O., Rotilio, G., Battistoni, A., and Djinovic-Carugo, K. (2004) Unique features of the sodC-encoded superoxide dismutase from *Mycobacterium tubercolosis*, a fully functional copper-containing enzyme lacking zinc in the active site. *J. Biol. Chem. 279*, 33447–33455.
12. Pesce, A., Capasso, C., Battistoni, A., Folcarelli, S., Rotilio, G., Desideri, A., and Bolognesi, M. (1997) Unique structural features of the monomeric Cu,Zn superoxide dismutase from *Escherichia coli*, revealed by X-ray crystallography. *J. Mol. Biol. 274*, 408–420.
13. Bordo, D., Matak, D., Djinovic-Carugo, K., Rosano, C., Pesce, A., Bolognesi, M., Stroppolo, M. E., Falconi, M., Battistoni, A., and Desideri, A. (1999) Evolutionary constraints for dimer formation in prokaryotic Cu,Zn superoxide dismutase. *J. Mol. Biol. 285*, 283.
14. Bourne, Y., Redford, S. M., Steinman, H. M., Lepock, J. R., Tainer, J. A., and Getzoff, E. D. (1996) Novel dimeric interface and electrostatic recognition in bacterial Cu,Zn superoxide dismutase. *Proc. Natl. Acad. Sci. U.S.A. 93*, 12774–12779.
15. Pesce, A., Battistoni, A., Stroppolo, M. E., Polizio, F., Nardini, M., Kroll, J. S., Langford, P. R., O'Neil, J. D., Sette, M., Desideri, A., and Bolognesi, M. (2000) Functional and crystallographic characterization of *Salmonella typhimurium* Cu,Zn superoxide dismutase coded by the sodCI virulence gene. *J. Mol. Biol. 302*, 465–478.
16. Gabbianelli, R., D'Orazio, M., Pacello, F., O'Neill, P., Nicolini, L., Rotilio, G., and Battistoni, A. (2004) Distinctive functional features in prokaryotic and eukaryotic Cu,Zn superoxide dismutases. *Biol. Chem. 385*, 749–754.
17. Ammendola, S., Pasquali, P., Pacello, F., Rotilio, G., Castor, M., Libby, S., Figueroa-Bossi, N., Bossi, L., Fang, F. C., and Battistoni, A. (2008) Regulatory and structural differences in the Cu,Zn-superoxide dismutase of *Salmonella enterica* and their significance for virulence. *J. Biol. Chem. 283*, 13688–99.
18. Banci, L., Bertini, I., Cramaro, F., Del Conte, R., Rosato, A., and Viezzoli, M. S. (2000) Backbone dynamics of human Cu, Zn superoxide dismustase and of its monomeric F50/EG51E/E133Q mutant: The influence of dimerization on mobility and function. *Biochemistry 39*, 9108–9118.
19. Arnesano, F., Banci, L., and Piccioli, M. (2006) NMR structures of paramagnetic metalloproteins. *Q. Rev. Biophys. 38*, 167–219.
20. Bermel, W., Bertini, I., Felli, I. C., Piccioli, M., and Pierattelli, R. (2006) ^{13}C-Detected *protonless* NMR spectroscopy of proteins in solution. *Prog. NMR Spectrosc. 48*, 25–45.
21. Caillet-Saguy, C., Delepierre, M., Lecroisey, A., Bertini, I., Piccioli, M., and Turano, P. (2006) Direct detected ^{13}C NMR to investigate the iron(III) hemophore HasA. *J. Am. Chem. Soc. 128*, 150–158.
22. Jiménez, B., Mori, M., Battistoni, A., Sette, M., and Piccioli, M. (2007) NMR assignment of reduced form of copper, zinc superoxide dismutase from *Salmonella enterica*. *Biomol. NMR Assignments 1*, 65–68.
23. Cole, R., and Loria, J. P. (2003) FAST-Modelfree: A program for rapid automated analysis of solution NMR spin-relaxation data. *J. Biomol. NMR 26*, 203–213.
24. Dalvit, C. (1996) Homonuclear 1D and 2D NMR experiments for the observation of solvent-solute interactions. *J. Magn. Reson., Ser. B 112*, 282–288.
25. Herrmann, T., Güntert, P., and Wüthrich, K. (2002) Protein NMR structure determination with automated NOE-identification in the NOESY spectra using the new software ATNOS. *J. Biomol. NMR 24*, 171–189.

26. Wishart, D. S., and Sykes, B. D. (1994) The ^{13}C chemical shift index: A simple method for the identification of protein secondary structure using ^{13}C chemical shift data. *J. Biomol. NMR 4*, 171–180.
27. Cordier, F., and Grzesiek, S. (1999) Direct observation of hydrogen bonds in proteins by interresidue $^{3h}J_{NC'}$ scalar couplings. *J. Am. Chem. Soc. 121*, 1601–1602.
28. Banci, L., Bertini, I., Cramaro, F., Del Conte, R., and Viezzoli, M. S. (2002) The solution structure of reduced dimeric copper zinc SOD: The structural effects of dimerization. *Eur. J. Biochem. 269*, 1905–1915.
29. Bertini, I., Duma, L., Felli, I. C., Fey, M., Luchinat, C., Pierattelli, R., and Vasos, P. R. (2004) A heteronuclear direct detection NMR experiment for protein backbone assignment. *Angew. Chem., Int. Ed. 43*, 2257–2259.
30. Bertini, I., Jiménez, B., Piccioli, M., and Poggi, L. (2005) Asymmetry in ^{13}C-^{13}C COSY spectra identifies geometry in paramagnetic proteins. *J. Am. Chem. Soc. 127*, 12216–12217.
31. Balayssac, S., Bertini, I., Luchinat, C., Parigi, G., and Piccioli, M. (2006) ^{13}C direct detected NMR increases the detectability of residual dipolar couplings. *J. Am. Chem. Soc. 128*, 15042–15043.
32. Bermel, W., Bertini, I., Felli, I. C., Lee, Y.-M., Luchinat, C., and Pierattelli, R. (2006) Protonless NMR experiments for sequence-specific assignment of backbone nuclei in unfolded proteins. *J. Am. Chem. Soc. 128*, 3918–3919.
33. Bertini, I., Capozzi, F., Luchinat, C., Piccioli, M., and Viezzoli, M. S. (1991) Assignment of active site protons in the 1H NMR spectrum of reduced human Cu,Zn superoxide dismutase. *Eur. J. Biochem. 197*, 691–697.
34. Folcarelli, S., Battistoni, A., Falconi, M., O'Neill, P., Rotilio, G., and Desideri, A. (1998) Conserved enzyme-substrate electrostatic attraction in prokaryotic Cu,Zn superoxide dismutase. *Biochem. Biophys. Res. Commun. 244*, 908–911.
35. Tainer, J. A., Getzoff, E. D., Beem, K. M., Richardson, J. S., and Richardson, D. C. (1982) Determination and analysis of 2 Å structure of copper zinc superoxide dismutase. *J. Mol. Biol. 160*, 181–217.
36. Djinovic, K., Battistoni, A., Carri, M. T., Polticelli, F., Desideri, A., Rotilio, G., Coda, A., and Bolognesi, M. (1994) Crystal structure of cyanide-inhibited *Xenopus laevis* Cu,Zn superoxide dismutase at 98 K. *FEBS Lett. 349*, 93–98.
37. Battistoni, A., Folcarelli, S., Rotilio, G., Capasso, C., Pesce, A., Bolognesi, M., and Desideri, A. (1996) Crystallization and preliminary X-ray analysis of the monomeric Cu,Zn superoxide dismutase from *Escherichia coli*. *Protein Sci. 5*, 2125–2127.
38. Banci, L., Bertini, I., Cantini, F., D'Onofrio, M., and Viezzoli, M. S. (2002) Structure and dynamics of copper-free SOD: The protein before binding copper. *Protein Sci. 11*, 2479–2492.
39. Djinovic, C. K., Battistoni, A., Carrì, M., Polticelli, F., Desideri, A., Rotilio, G., Coda, A., Wilson, K., and Bolognesi, M. (1996) Three-dimensional structure of *Xenopus laevis* Cu,ZnSOD *b* determined by X-ray crystallography at 1.5 Å resolution. *Acta Crystallogr. D52*, 176–188.
40. Stroppolo, M. E., Nuzzo, S., Pesce, A., Rosano, C., Battistoni, A., Bolognesi, M., Mobilio, S., and Desideri, A. (1998) On the coordination and oxidation states of the active-site copper ion in prokaryotic Cu,Zn superoxide dismutase. *Biochem. Biophys. Res. Commun. 249*, 579–582.
41. Bertini, I., Dalvit, C., Huber, J. G., Luchinat, C., and Piccioli, M. (1997) ePHOGSY experiment on a paramagnetic protein: Location of the catalytic water molecule in the heme crevice of the oxidized form of horse heart cytochrome c. *FEBS Lett. 415*, 45–48.
42. Getzoff, E. D., Cabelli, D. E., Fisher, C. L., Parge, H. E., Viezzoli, M. S., Banci, L., and Hallewell, R. A. (1992) Faster superoxide dismutase mutants designed by enhancing electrostatic guidance. *Nature 358*, 347–351.
43. Falconi, M., Stroppolo, M. E., Cioni, P., Strambini, G., Sergi, A., Ferrario, M., and Desideri, A. (2001) Dynamics-function correlation in Cu,Zn superoxide dismutase: A spectroscopic and molecular duynamics simulation study. *Biophys. J. 80*, 2556–2567.
44. Arnesano, F., Banci, L., Bertini, I., Martinelli, M., Furukawa, Y., and O'Halloran, T. V. (2004) The unusually stable quaternary structure of human SOD1 is controlled by both metal occupancy and disulfide status. *J. Biol. Chem. 279*, 47998–48003.
45. Bertini, I., Luchinat, C., Ming, L.-J., Piccioli, M., Sola, M., and Valentine, J. S. (1992) Two-dimensional ^1H-NMR studies of the paramagnetic metalloenzyme copper-nickel superoxide dismutase. *Inorg. Chem. 31*, 4433–4435.
46. Stroppolo, M. E., Pesce, A., D'Orazio, M., O'Neill, P., Bordo, D., Rosano, C., Milani, M., Battistoni, A., Bolognesi, M., and Desideri, A. (2001) Single mutations at the subunit interface modulate copper reactivity in Photobacterium leiognathi Cu,Zn superoxide dismutase. *J. Mol. Biol. 308*, 555–563.
47. Forest, E., Langford, P. R., Kroll, J. S., and Getzoff, E. D. (2000) Cu,Zn superoxide dismutase structure from a microbial pathogen establishes a class with a conserved dimer interface. *J. Mol. Biol. 296*, 145–153.
48. Fang, F. C., DeGroote, M. A., Foster, J. W., Bäumler, A. J., Ochsner, U., Testerman, T., Bearson, S., Giàrd, J. C., Xu, Y., CAmpbell, G., and Laessig, T. (1999) Virulent *Salmonella typhimurium* has two periplasmic Cu, Zn-superoxide dismutases. *Proc. Natl. Acad. Sci. U.S.A. 96*, 7502–7507.
49. Koshunov, S. S., and Imlay, J. A. (2006) Detection and quantification of superoxide formed within the periplasm of *Escherichia coli*. *J. Bacteriol. 188*, 6326–6334.

BI801252E

Chapter 6

6. Towards structural dynamics: fluctuations in proteins monitored by chemical shift modulations and direct detection of C'N multiple-quantum relaxation rates

In biological systems the function of biomacromolecules is exquisitely dependent on their spatial and temporal fluctuations and/or modifications. Several biological processes are based on transduction of information through conformational changes in proteins and nucleic acids associated with folding and assembly, ligand binding and molecular recognition, and catalysis. A central problem in understanding biological processes at a molecular level is the elucidation of how the active conformation of biomacromolecules is achieved on time scales necessary for function. Technological developments have revolutionized the range of spectroscopic and other approaches available for the study of dynamic processes in biomacromolecules; however, NMR spectroscopy has a unique capacity to investigate dynamic properties of molecules over a range of different time scales with atomic resolution in both solution and solid states.

To date, a huge number of three dimensional structures of proteins and nucleic acids have been determined by X-ray diffraction and NMR spectroscopy. While this effort has provided insights into protein conformation, chemical composition of reactive sites and biomolecular interfaces, it has also become clear that knowledge of the static structure alone is insufficient to explain how the biomolecules precisely work. In fact, proteins exist as complex ensembles of conformations that are continuously interconverting due to thermal fluctuations. Only a subset of these conformations is competent for any particular function, so the observed functional properties of the protein are a manifestation of the functional properties of each conformational state, of the populations of these states (thermodynamics of the ensemble), and of the rates of interconversion between the different conformations (kinetics of the ensemble). Dynamics is the process of interconversion between conformational states. Therefore, studying protein dynamics we can obtain information about the time scales of motions

and the population distribution of conformational states. The energy barriers separating different conformations of a protein can vary dramatically, so interconversion between states can be as fast as a few picoseconds (for librations and rotations of small groups) and as slow as many seconds (for large conformational rearrangements such as unfolding).

The characterization of the dynamics of biomolecules plays a central role in the study of biological systems. The determination of dynamic properties can provide information on the activity of the molecule, and on the change of activity upon interaction with other molecules.

Since the first NMR experiment was performed, researchers set out to discover the factors leading to the particular peak intensities and linewidths observed and had begun to address the various aspects of NMR relaxation phenomena [1,2]. NMR relaxation is caused by fluctuations of the Hamiltonian describing the relaxing spin system and the major source of this fluctuation is the reorientation of the interacting spins with respect to each other and to the external magnetic field. The time scale of a dynamic process that can be characterized by spin relaxation methods depends directly on the magnitude of the variation in the Hamiltonian modulated by the dynamics process. In other words, NMR spin relaxation can provide information on motions occurring in molecules. Initially, Bloembergen et al. introduced the concept of motional narrowing, the effect of molecular motion on observed linewidths to explain the narrow lines observed in liquids as compared to solids. This concept was later applied to study internal nuclear motion in solids, an application which is useful in the determination of internal motions in proteins [3]. Expressions for the longitudinal (T_1) relaxation, transverse (T_2) relaxation and the Nuclear Overhauser Effect (NOE) were developed by Abragam and Pound [4], and Solomon [1] for two-spin systems that were assumed to reorient their inter-nuclear vector randomly and isotropically.

During the years, it became clear that a more informative interpretation of relaxation information was desired beyond the simple reporting of relaxation times. Indeed, an explicit connection between the measured relaxation parameters and the dynamics parameters of the system under study was desired.

In 1982, such a link was provided by the so-called ''Model-Free'' approach of Lipari and Szabo [5], which allows the unique information on ''fast'' internal motions (faster than the overall tumbling, i.e. nano- to picoseconds dynamics) to be completely characterized by two motional model independent parameters at each site. These parameters are the generalized order parameter, S^2, which measures the spatial restriction of the internal motion, and the effective local correlation time, τ_e, which measures the rate of the internal or local motion. Coupling the numerical values of S^2 and τ_e with a physical model allows one to generate a physical picture of the motion.

The basic premise of this formalism is that the internal motions of bond vectors in proteins are independent of the overall rotational diffusion of the molecule. In addition, the rotational diffusion of the molecule influences each bond vector identically (for isotropic rotation) or in a manner that is related through the relative orientations of the bond vectors in the molecule (for non-isotropic rotation), whereas the internal motions of any two bond vectors are independent of each other or at least unrelated in any predictable way. However, this assumptions give limitation on the potential accuracy of the method and they don't provide a detailed physical picture of the internal motions that can then be related to the chemical mechanism, binding interactions or other physical or spectroscopic properties of the protein.

Slow dynamics have also been shown to correlate with biological activity [6,7]. There is growing interest in slow phenomena such as fluctuations of hydrogen bonds and rearrangements of loops, since such motions are thought to be responsible for the specificity of phenomena like protein interaction with ligands, protein folding, protein-protein and protein-nucleic acid interaction.

Chemical exchange is a phenomenon in NMR spectroscopy that provides information on conformational and chemical kinetic process occurring on μs-s time scales. Chemical exchange is a processes that modulate isotropic chemical shifts by altering the magnetic environments of spins. In NMR, they can be probed through transverse relaxation measurements. The main experiments for quantifying chemical exchange are longitudinal magnetization exchange [8], line shape analysis [9], CPMG relaxation dispersion [10] and $R_{1\rho}$ relaxation dispersion [10]. However, the complexity of the dynamics prevents a satisfactory characterization of the motions, and further

experimental data are required to obtain a more complete picture. All of these methods rely on extracting the effect of exchange from the autocorrelation rates of the nuclei studied. In large biological systems, such as proteins, the residues are likely to be subject to complicated changes of conformations, so that the information provided by autocorrelation measurements alone is insufficient to enable a clear characterization of these processes.

The understanding of the dynamics process is often achieved by measuring not only the relaxation due to single mechanisms but also the interference between two relaxation mechanisms. Cross correlated relaxation rates, commonly referred as cross correlation rates (CCR), arise from the interference of two relaxation mechanisms which are modulated by the same correlation time. In presence of internal reorientations, CCR contains information on the correlate motions of the tensors which describe the relaxation mechanisms in the three dimensional space. Cross-correlation rates can provide a more refined picture and, in some cases, detailed hypotheses can be made about the precise nature of the motions and sophisticated models can provide a more complete description of dynamic processes [11,12].

Our interest here is to study how correlated modulations of the chemical shifts of nuclei involved in multiple quantum coherences can lead to cross-correlation effects. The measured cross-correlation rates are not only sensitive to the presence of slow motions, but also give information on the extent of correlation between the modulations of the chemical shifts of the two nuclei. If two spins are affected by the same chemical exchange kinetic process, then the chemical shift changes for the two spins resulting from transitions between sites will be correlated. This correlation gives rise to exchange effects that can either broaden or narrow resonance line shapes for multiple quantum coherences [13].

References

[1] I. Solomon, Phys. Rev. 99 (1955), 559-565.

[2] A.G. Redfield, Adv. Magn. Reson. 1 (1965), 1-32.

[3] A. Mittermaier, L.E. Kay, Science 312 (2006), 224-228.

[4] A. Abragam, R.V. Pound, Phys. Rev. 92 (1953), 953.

[5] G. Lipari, A. Szabo, J. Am. Chem. Soc. 104 (1982), 4559-4570.

[6] A.P. Kalverda, M. Ubbink, G. Gilardi, S.S. Wijmenga, A. Crawford, L.J. Jeuken, G.W. Canters, Biochemistry 38 (1999), 12690-12697.

[7] F.A.A. Mulder, N.R. Skrynnikov, B. Hon, F.W. Dahlquist, L.E. Kay, J. Am. Chem. Soc. 123 (2001), 967-975.

[8] J.J. Led, H. Gesmar, F. Abilgaard, In: (176 ed.) (1989), pp. 311-329.

[9] J. Sandstrom, In: Dynamic NMR Spectroscopy, Academic Press, London (1982).

[10] D.G. Davis, M.E. Perlman, R.E. London, J. Magn. Reson. Ser. B 104 (1994), 266-275.

[11] D. Frueh, Prog. NMR Spectrosc. 41 (2002), 305-324.

[12] C. Perazzolo, J. Wist, K. Loth, L. Poggi, S.W. Homans, G. Bodenhausen, J. Biomol. NMR 33 (2005), 233-242.

[13] D. Frueh, J.R. Tolman, G. Bodenhausen, C. Zwahlen, J Am Chem Soc. 123 (2001), 4810-4816.

Copy of the manuscript submitted to Journal of American Chemical Society

Towards structural dynamics: fluctuations in proteins monitored by chemical shift modulations and direct detection of C'N multiple-quantum relaxation rates

Mirko Mori[a], Fatiha Kateb[b,c], Geoffrey Bodenhausen[b], Mario Piccioli[a], and Daniel Abergel[b*]

[a] *Magnetic Resonance Center (CERM) and Department of Chemistry, University of Florence, Via L. Sacconi 3, 50019 Sesto Fiorentino, Italy*

[b] *Laboratoire des Biomolécules, associé au CNRS, Département de Chimie, Ecole Normale Supérieure, 24, rue Lhomond 24, 75231 Paris cedex 05, France*

[c] present address: *Lehrstuhl für Biomolekulare NMR-Spektroskopie, Departement Chemie, Lichtenbergstr. 4 D-85747 Garching, Germany*

** Daniel.Abergel@ens.fr*

Abstract

Multiple quantum relaxation in proteins reveals unexpected relationships between correlated or anti-correlated conformational backbone dynamics in α helices or β sheets. The contributions of conformational exchange to the relaxation rates of C'N coherences (i.e., double- and zero-quantum coherences involving backbone carbonyl ^{13}C' and neighbouring amide ^{15}N nuclei) depend on the kinetics of slow exchange processes, as well as on the populations of the conformations and chemical shift differences of ^{13}C' and ^{15}N nuclei. The relaxation rates of C'N coherences, which reflect concerted fluctuations due to slow chemical shift modulations (CSM), were determined by direct ^{13}C detection in diamagnetic and paramagnetic proteins. In well-folded proteins such as Lanthanide-substituted Calbindin (CaLnCb), Copper-Zinc Superoxide Dismutase (Cu,ZnSOD) and Matrix Metalloproteinase (MMP12), slow conformational exchange occurs along the entire backbone. Our observations demonstrate that relaxation rates of C'N coherences arising from slow backbone dynamics have positive signs (characteristic of correlated fluctuations) in β sheets, and negative signs (characteristic of anti-correlated fluctuations) in α helices. This extends the prospects of *structure-dynamics* relationships to slow time scales that are relevant for protein function and enzymatic activity.

Introduction

In recent years, there has been growing evidence that structural flexibility plays a key role for protein function. Many biochemical events such as enzymatic reactions, the formation or disruption of hydrogen bonds, the alteration of dihedral angles, and the reorientation of aromatic rings, occur on slow micro- to millisecond time scales. The functional relevance of protein dynamics has been amply demonstrated, and its study by NMR has been one of our main goals. Beyond the celebrated *structure-function* relationships, several groups have proposed to investigate the existence of *structure-dynamics* relationships. We like to refer to "structural dynamics" in this context, since the expression "structural biology" has an unduly static connotation. Several empirical and theoretical attempts have been made to relate experimental observations of protein dynamics to molecular structure. In most studies, internal mobility is described in terms of rapid fluctuations about an average structure. These fluctuations are often modelled using local harmonic potentials, as in Normal Mode Analysis (NMA)[1], Gaussian Network Models (GNM)[2], Networks of Coupled Rotators (NCR)[3], and other methods. These approaches are mostly relevant to describe fast (sub-nanosecond) internal dynamics as reflected in relaxation rates (R_1, R_2 and Overhauser effects) of isolated ^{15}N or ^{13}C nuclei, although quantities such as order parameters and conformational entropy are not associated with any specific time scale of internal dynamics.

The situation is clearly different for conformational exchange, where NMR can provide insight into slow processes that occur in a μs-ms range. The contributions of conformational exchange processes to relaxation rates reflect the kinetic rate, and the populations and chemical shifts of the nuclei in these conformations. Therefore, there is no simple relationship between NMR relaxation rates due to conformational exchange

and protein structure, even if there may be some correlations between relaxation rates and various structural motifs such as α helices and β sheets. Structural fluctuations in the protein backbone may lead to observable chemical shift modulations (CSM), the magnitudes and signs of which are not easy to predict.

In this work, these fundamental questions are addressed for a few well-structured proteins with different folds. Slow backbone fluctuations were studied by measuring isotropic chemical shift modulations (CSM's) of carbonyl/amide C'N coherences, using both novel direct ^{13}C detection and known indirect ^{1}H detection techniques. Direct ^{13}C detection *('protonless')* experiments are to be preferred when ^{1}H signals suffer from (possibly paramagnetic) line-broadening[4,5]. Even observables involving protons, such as $H^{N}N$ or $H^{\alpha}C^{\alpha}$ residual dipolar couplings (RDC's), can be obtained using pulse sequences without detecting or exciting the protons directly[6]. Here, we extend the protonless methodology to dynamic studies. Direct detection of ^{13}C nuclei offers an alternative route to the measurement of relaxation rates of C'N multiple-quantum coherences,[7] which may contribute to a more complete description of backbone dynamics in proteins[8] as a complementary tool to conventional techniques that focus on the relaxation of isolated ^{15}N and ^{13}C nuclei[9]. Protonless experiments avoid losses of information caused by fast relaxation or exchange broadening of ^{1}H signals and may therefore offer better sensitivity, particularly in the presence of paramagnetic ions.

The observations presented in this work demonstrate that conformational exchange is not restricted to some regions of the backbone; rather, it is a phenomenon that extends across the entire backbone of structured proteins. Moreover, these results allowed us to discover a surprisingly simple correlation between the *type of secondary*

structure (α helices or β sheets) and the *sign* of the CSM contributions, which is negative in α helices and positive in β sheets. These observations represent a significant step in understanding structure/dynamics relationships. A tentative rationalization will be proposed, based on published theoretical analyses.

Results

Direct ^{13}C detection. The 75-residue protein Calbindin D_{9k} in its calcium-loaded form (Ca_2Cb) was used to test the protonless sequence of Fig. 1 (see *Methods* for details). This protein comprises two EF-hand domains and can bind two Ca^{2+} ions[10]. The assignment, structure and dynamics of Calbindin are well known[9-12]. One of the two calcium ions can be easily replaced by a metal ion of the lanthanide series[13]. Such substitutions lead to systems with similar structures but different spectroscopic properties, depending on the lanthanide ion[11,14]. Here, we studied diamagnetic La^{3+} and paramagnetic Ce^{3+}-substituted derivatives (CaLaCb and CaCeCb). For each of these systems, the difference ΔR(C'N) of the relaxation rates of zero- and double-quantum coherences involving $^{13}C'$ and ^{15}N nuclei belonging to the same peptide bond (see *Methods*) was measured twice, either by traditional HNCO-type experiments (using coherence transfer from and to ^{1}H nuclei), or by protonless ^{13}C detected CON-type sequences (Fig.1). As shown in Fig. 2, the results obtained by the two methods are in good agreement and provide reliable measurements of multiple-quantum relaxation rates ΔR(C'N).

For the diamagnetic CaLaCb form of Calbindin, the ΔR(C'N) rates of 57 of the 74 amide bonds could be measured with both approaches. Two further rates could only be determined with conventional proton detection, whilst 6 additional rates could only be measured using protonless experiments, thus giving a total of 65 rates. This shows that

the two experiments are complementary. Even in the absence of paramagnetic centers, protonless experiments can provide more rates than proton-detected methods.

In the case of the paramagnetic CaCeCb form of Calbindin, there are 9 amide bonds for which the rates ΔR(C'N) could *only* be determined by protonless experiments, and 54 signals that could be detected in both ^{1}H and ^{13}C experiments. This could be expected, as signal losses due to paramagnetic relaxation are more dramatic for protons because of their large gyromagnetic ratios[15]. These observations demonstrate the interest of direct ^{13}C detection.

If one neglects relaxation, ^{1}H-observed experiments should have a sensitivity gain of $(\gamma_H/\gamma_C)^{5/2} = 32$ compared to experiments that start and end with ^{13}C nuclei. However, relaxation is the main cause of intensity losses in our case. Magnetization can be lost through transverse ^{1}H relaxation during the INEPT transfer steps of the HNCO sequence, and even if these fixed delays represent only roughly 20% of the total duration of the sequence, their removal dramatically reduces relaxation losses. Furthermore, INEPT and inverse INEPT steps require a large number of ^{1}H and ^{15}N pulses, which also contribute to signal losses because of pulse imperfections. Coherence transfer steps in protonless ^{13}C-detected experiments are not affected by transverse ^{1}H relaxation, as no proton coherences are involved. This issue is illustrated in Fig. 3, which shows selected traces taken from HNCO and CON experiments in diamagnetic CaLaCb and paramagnetic CaCeCb. For diamagnetic systems, the S/N ratio is better in ^{1}H-detected than in ^{13}C-detected experiments by a factor of about three (Figs. 3A-B). In the paramagnetic CaCeCb system, transverse proton relaxation of metal-binding residues can be so fast that the corresponding H^N signals are not visible in the HNCO-type experiment. In contrast, in ^{13}C-detected experiments, ΔR(C'N) rates can be

determined even for metal-bound residues. This is illustrated in Fig. 3C for Asn 56, which is close to Ce^{3+}. No such signal can be observed in HNCO-type experiments. Note that the two ways of measuring ΔR(C'N) rates are complementary, since the dispersion of the signals in C'N 2D spectra is comparable, if not better, than in H^NN spectra.

Slow backbone fluctuations in proteins and structure-dynamics relationships

The above techniques were applied to study slow backbone dynamics of an ensemble of structurally different proteins: Calbindin (Cb), Cu,Zn-Superoxide Dismutase (henceforth denoted Cu,ZnSOD) and Matrix Metalloproteinase 12 (MMP12).

Measured ΔR(C'N) backbone rates in Calbindin are mostly negative, -13 s^{-1} < ΔR(C'N) < -2 s^{-1}. Following the procedure described in previous work[8], $R^{CSM/CSM}$ rates were extracted from the measured ΔR(C'N) rates. For each sample, the mean rates and their standard deviation were calculated as described in the *Methods* section. One thus finds $< R^{CSM/CSM} >$ = -2.39 s^{-1} and σ = 1.28 s^{-1} for CaLaCb. The results reported in Fig. 4A show that significant $R^{CSM/CSM}$ contributions exist across the entire backbone of the protein. Using a terminology introduced previously,[7,8,16,17] these negative $R^{CSM/CSM}$ rates indicate that mostly anti-correlated modulations are observed throughout the Calbindin backbone. In addition, significant deviations of $R^{CSM/CSM}$ rates from their average occur for 16 residues. Remarkably, 14 out of the 16 outlying residues are located in the calcium binding loops I and II, in the linker region, and in the C- and N-terminal ends. Moreover, EF-hand loops I (residues 14-27) and II (residues 54-65) feature different behaviours. In loop I, deviations $| R^{CSM/CSM} - < R^{CSM/CSM} > |$ that are much larger than the standard deviation are observed for five residues (Lys 16, Glu 17, Gly 18, Gln 22, Ser

24), indicating more pronounced exchange contributions. As shown in Fig. 4A, four of the $R^{CSM/CSM}$ rates are smaller (i.e., less negative) and one is larger (more negative). Alternatively, among the three residues in loop II (Asn 56, Asp 58, Ser 62) that have $R^{CSM/CSM}$ rates that deviate from the average, two (Asn 56 and Asp 58) have rates that are larger in magnitude (more negative). In addition, Glu 65 also appears to exhibit a lower rate, although our criteria led us to discard it from further analysis. Interestingly, the $R^{CSM/CSM}$ rates of Glu 17 and Gln 22 in binding loop I, which bind the metal ion via their backbone carbonyl oxygens, lie above the average (depicted in red in Fig. 4A). The $R^{CSM/CSM}$ rates of Asn 56 and Asp 58 in loop II (shown in blue in Fig. 4A), which bind the metal ion via their side-chain carboxylate groups, have the largest (most negative) rates. These findings show that exchange in Calbindin, as reflected in the values of the $R^{CSM/CSM}$ rates, increases in binding loops, and suggest that the correlation or anti-correlation of the chemical shift fluctuations of the ^{13}C' and ^{15}N nuclei may be related to the way residues are bound to the metal, namely *via* backbone carbonyl groups (loop I) or *via* side-chain carboxylate groups (loop II).

Significant deviations of the $R^{CSM/CSM}$ rates from their average were also observed in the linker region (Pro 37, Gly 42 and Thr 45), as well as in the C- and N-terminal ends. These regions are also known to undergo internal dynamics on a sub-nanosecond time scale, as demonstrated by ^{15}N relaxation measurements[9]. However, no significant exchange contributions had been detected before we determined the $R^{CSM/CSM}$ rates. This confirms the ability of backbone C'N coherences to reveal fluctuations of the backbone structure through complementary ^{13}C direct and ^{1}H indirect detection methods.

These results corroborate previous observations[8,16] that proteins may undergo

conformational exchange across their entire backbone. The presence of increased exchange in more loosely structured regions such as loops and linkers raises the possibility of a relationship between secondary structure elements and dynamics in the μs-ms time scale. This point was further investigated by measuring $R^{CSM/CSM}$ rates in Cu,ZnSOD and MMP12 by protonless direct ^{13}C detection.

The protein Cu,ZnSOD (PDB entry 2K4W) is a monomeric 154 amino-acid enzyme, isolated from *Salmonella enterica*[18], and comprises an eight-stranded greek-key β-barrel, two small α-helices, and a loop region stabilized by a Zn^{2+} ion. The enzymatic cavity is defined by an electrostatic loop, which contains a copper ion that can jump between oxidation states $Cu^{2+/+}$ and is responsible for the catalytic activity. The three-dimensional structure of this enzyme has been determined in solution,[19] and the complete NMR assignment is available[20]. Crystal structures of many homologous enzymes are also available[21].

The $R^{CSM/CSM}$ rates obtained for Cu,ZnSOD are shown in Fig. 4B. The average trimmed rate was (see *Methods*) $< R^{CSM/CSM} > = -0.86\ s^{-1}$ with a standard deviation $\sigma = 3.69\ s^{-1}$. The rates are scattered across a wide range. There is a clear-cut correlation between the rates obtained from the experiments and the secondary structure (Fig. 4B). Indeed, for the 57 residues located in β sheets, one finds $< R^{CSM/CSM} >_\beta = +1.94 \pm 1.90\ s^{-1}$, whereas the average rate observed for the 9 residues in the short α helices is $<R^{CSM/CSM}>_\alpha = -6.13 \pm 2.24\ s^{-1}$. In the loop regions the average rate is $< R^{CSM/CSM} >_l = -2.16 \pm 3.71\ s^{-1}$. Therefore, in the β-barrel regions of Cu,ZnSOD, the $R^{CSM/CSM}$ rates are positive and lie in the range $0 < R^{CSM/CSM} < +4\ s^{-1}$, whereas in the α-helical fragment they are negative. We shall see below that the positive rates $R^{CSM/CSM}$ typical of β sheets are characteristic of correlated fluctuations of chemical shifts, while the negative rates

found α helices reveal anti-correlated fluctuations. In loop regions, the $R^{CSM/CSM}$ rates are intermediate between those measured in α-helices and β-barrels, and characterized by larger standard deviations than those observed in well-defined secondary structure elements.

Further evidence of a relationship between secondary structure elements and $R^{CSM/CSM}$ rates was provided by experiments performed on MMP12 (Fig. 4C). Matrix metalloproteinases are composed of a catalytic domain, and of a hemopexin-like domain[22]. The catalytic domain adopts a typical MMP fold and binds one "structural" Zn^{2+} ion as well as three Ca^{2+} ions, in addition to the catalytic Zn^{2+} ion[23]. The 3D fold is composed of five β-sheets flanked by the Zn^{2+} binding loop, and two long α-helices that also include the active site. The C-terminal region comprises a long loop and a small α-helix. Structures determined by solution NMR[24] and by X-ray diffraction [25] (PDB entries 1Y93 and 1RMZ) of this protein are available. As this enzyme contains both α-helices and β-sheets, it seemed an appropriate system to complement our observations on the α-helices of Calbindin and the β barrel protein Cu,ZnSOD.

The average $R^{CSM/CSM}$ rates calculated for each secondary structure in MMP12 indicate a similar pattern, and therefore corroborate our previous observations. Indeed, the β-sheet features correlated motions, with $< R^{CSM/CSM} >_\beta = +2.46 \pm 1.85$ s^{-1}, whereas the α helices show anti-correlated CSM motions with $< R^{CSM/CSM} >_\alpha = -5.04 \pm 2.29$ s^{-1}. Finally, as in the cases of Calbindin and Cu,ZnSOD, the behaviour of loop regions is characterized by negative $R^{CSM/CSM}$ rates with a much larger scatter $< R^{CSM/CSM} >_l = -3.24 \pm 3.49$ s^{-1}.

Discussion

The three proteins studied in this work have very different properties, in terms of

secondary structures, three-dimensional folds, biological function, and NMR relaxation rates. Nevertheless, our observations reveal a relationship between secondary structure elements and $R^{CSM/CSM}$ rates. The evidence shows that β-sheets and α helices are associated with correlated and anti-correlated chemical shift modulations. This relationship is striking when representing $R^{CSM/CSM}$ rates against backbone ψ angles for the three proteins (Fig. 5). The range ψ = -45° ± 25°, typical of α-helical structures, corresponds to negative $R^{CSM/CSM}$ rates. Angles in the range ψ = +140° ± 30° are characteristic of β-sheets and correspond to positive $R^{CSM/CSM}$ rates. The empirical relationship between $R^{CSM/CSM}$ rates and the type of secondary structure element thus turns out to be a general feature. The contribution to multiple quantum relaxation rates of C'N coherences due to chemical shift modulations arising from fast two-site exchange is given by[26]:

$$R^{CSM/CSM} = 2p_a p_b \tau_{ex} \Delta\omega_{C'} \Delta\omega_N \qquad (1)$$

where $\tau_{ex} = 1/(k_{AB} + k_{BA})$, $\Delta\omega_{C'}$ and $\Delta\omega_N$ are the changes of the isotropic chemical shifts experienced by C' and N nuclei when hopping from one site to another, and p_a and p_b are the populations of the two conformations.[7] The sign of the rate $R^{CSM/CSM}$ depends on the relative signs of $\Delta\omega_{C'}$ and $\Delta\omega_N$. Therefore, valuable insight can be gained by correlating chemical shift *variations* with specific secondary structure elements. To our knowledge, only relationships between 'static' C' and N chemical shifts (as opposed to chemical shift *modulations*) and backbone secondary structure have been addressed so far.[27] It is clear that backbone motions, i.e., modulations of ϕ, ψ angles accompanied by the making and breaking of hydrogen bonds, should induce modulations of isotropic chemical shifts. To rationalize our observations, one needs information not only about the dihedral angles that are affected by the dynamic processes, but also about the

relationship between structures and chemical shifts. Many different factors, such as hydrogen bonding, backbone conformation, and side-chain conformation of neighbouring residues, all contribute to ^{15}N chemical shifts in proteins.[28] Backbone ^{13}C chemical shifts are apparently more sensitive to side-chain conformations than to the neighbouring residues.[29]

A recent model developed by Wingreen and co-workers aims at predicting characteristic motions in helices and sheets.[30,31] According to these authors, β sheets undergo twisting or bending about an axis located in the plane of the sheet. In this case, the different strands that constitute the sheet should undergo displacements with respect to one another, without deformations of the strands themselves, therefore leaving all backbone ψ angles unchanged. On the other hand, inter-strand distances are modulated during this kind of motion, so that a weakening of hydrogen bonds should be expected. Density functional theory (DFT) calculations by Xu and Case[29] predict variations of ^{15}N and ^{13}C chemical shifts in model peptides that adopt β sheet conformations. Their predictions show that both amide ^{15}N and carbonyl ^{13}C chemical shifts move to higher frequencies when going from an isolated β strand to the central strand of a triple-stranded β sheet, due to modifications of hydrogen bond patterns and strengths. This would lead to a positive product $\Delta\omega_C \Delta\omega_N$, and therefore to a positive $R^{CSM/CSM}$ rate, in agreement with our observations. For double-stranded sheets, the calculated ^{13}C chemical shifts differences are again positive but small compared to a single-strand configuration, although they may lead to small negative $R^{CSM/CSM}$ rates.[29] This could explain some of the weak negative contributions to $R^{CSM/CSM}$ observed in β sheets.

In our observations in three proteins, 91 residues are located in β strands. Among them, 69 are associated with positive $R^{CSM/CSM}$, in accordance with the above

line of reasoning, whereas for 13 residues the CSM rates could not be measured. The remaining 9 residues have negative $R^{CSM/CSM}$ rates, 8 of which were observed in Cu,ZnSOD. These rates need to be analyzed on the basis of Cu,ZnSOD structure, and confronted with our hypothesis. The corresponding residues are Ser 18, Ile 19, Thr 27, Thr 34, Gly 46, His 48, Asp 93 and Tyr 148. Ser 18 and His 48 have rates $R^{CSM/CSM}$ = -0.37 ± 0.38 s^{-1} and -0.49 ± 0.51 s^{-1}. Considering the experimental errors, these rates are not significantly negative. Gly 46 belongs to the β strand 44-50, which is flanked by the β sheet 124-128. The partner residue of Gly 46 in this strand is His 128, which is situated at the end of the β sheet at the transition to the nearby loop, so that the hydrogen bond pattern is disrupted in the case of Gly 46. Therefore, basing our interpretation on the model calculations of Xu and Case[29] one should not expect clearly positive $R^{CSM/CSM}$ rates. The same reasoning can be applied to Thr 34. Residues Thr 27 and Asp 93 are located at the end of a β strand, and therefore fall into the same category. Finally, Ile 19 and Tyr 148 have unusual ψ angles, which may explain the negative $R^{CSM/CSM}$ rates. In the case of MMP12, the residue Ala 60, which is located at the end of a β strand, is also associated with a negative $R^{CSM/CSM}$ rate.

According to the description of Wingreen and co-workers, the main motional modes in α helices consist of bending and twisting motions about the average helix axis.[31] Such motions imply the occurrence of fluctuations of backbone ψ angles. The ^{15}N chemical shifts depend in a complex manner not only on the dihedral angles ψ(i-1), ϕ(i), and ψ(i) of the backbone, but also on the dihedral angle χ_1(i) of the side-chain. The ^{15}N chemical shifts are also sensitive to the side-chain dihedral angle χ_1(i-1) of the preceding residue through the nearest neighbour effect[32], and furthermore affected by hydrogen bonding.[29,33] On the other hand, carbonyl ^{13}C chemical shifts are mainly sensitive to

backbone torsion angles $\psi(i)$ and $\phi(i+1)$,[34] and therefore give reliable reports on secondary structure. Finally, the average backbone chemical shifts in α helices were found to be +1.7 ppm for ^{13}C and -1.7 ppm for ^{15}N nuclei with respect to random coil rates.[35] A transient deformation of a helical region towards a random-coil-like conformation should therefore be associated with $\Delta\omega_{C'}\Delta\omega_N < 0$. This is consistent with our experimental data, which show mostly negative $R^{CSM/CSM}$ rates in helical regions of the proteins.

Concluding remarks

An unexpected relationship was found to exist between slow conformational backbone dynamics and secondary structure elements in proteins. Slow backbone dynamics were identified through relaxation rates $R^{CSM/CSM}$ of C'N multiple quantum coherences. These rates provide more information than longitudinal and transverse relaxation rates of isolated ^{15}N nuclei, since the $R^{CSM/CSM}$ rates depend on the rates of interconversion between the conformations, on their populations, and on their ^{15}N and ^{13}C chemical shifts. Direct ^{13}C detection of $R^{CSM/CSM}$ rates proved to be valuable for both diamagnetic and paramagnetic proteins. Measurements of $R^{CSM/CSM}$ rates on three different folded proteins revealed the existence of extensive conformational exchange across the entire protein backbones. The relaxation rates of C'N coherences arising from slow backbone dynamics have positive signs (characteristic of correlated fluctuations of the chemical shifts of neighbouring ^{15}N and ^{13}C nuclei) in β sheets, and negative signs (characteristic of anti-correlated shift fluctuations) in α helices. These observations extend the prospect of *structure-dynamics* relationships to time scales on which protein function and enzymatic activity actually take place. In our view, this represents a significant step towards the development of functional dynamics.

Materials and Methods

Sample Preparation. Protein expression and purification of native, Ca^{2+} loaded Calbindin D_{9k} (Ca_2Cb), was performed as described elsewhere[36]. Lanthanide- (Ln^{3+}) substituted derivatives (CaLaCb and CaCeCb) were obtained following the established procedure[13]. Sample concentrations were comprised between 0.6 to 0.9 mM.

Uniformly ^{13}C-, ^{15}N- labeled Copper(I), Zinc(II)-Superoxide Dismutase was prepared and purified as described elsewhere[19]. The protein was concentrated to 1.0 mM and reduction was achieved by addition of a 0.10 M solution of sodium isoascorbate in phosphate buffer at pH 6.0.

The catalytic domain of MMP12, corresponding to the segment Gly-106–Gly-263, was cloned and expressed as reported in literature[37]. After refolding, the catalytic domain was inhibited by *N*-isobutyl-*N*-[4-methoxyphenylsulphonyl] glycyl hydroxamic acid (NNGH). The 0.9 mM sample of ^{13}C-, ^{15}N-labelled MMP12 had a pH adjusted to 7.2.

NMR experiments. All experiments were carried out on a Bruker Avance 700 spectrometer at 298 K. Proton-detected experiments were performed with a standard inverse-detection, triple-resonance probe (TXI), while a customized triple-resonance probe with a ^{13}C-selective inner coil was used for ^{13}C-detected experiments. Protonless CON-type multiple-quantum experiments were performed using the pulse sequence shown in Fig. 1. The 2D matrices consisted of 1024 and 128 points in the t_2 and t_1 dimensions. To determine auto- and cross-correlated relaxation rates, 128 and 512 scans were accumulated, respectively, with a 1.3 s relaxation delay. HNCO-type spectra were recorded using an established pulse sequence,[17] with 80 and 320 scans to reveal auto- and cross-peaks, respectively. Data processing was performed using the TOPSPIN

software.

Analysis of $R^{CSM/CSM}$ relaxation rates. The mean and standard deviation of the ΔR rates were calculated for the complete set of data for each system following a procedure described elsewhere[8]. Outliers, defined as rates that lie farther that 1.5 standard deviations from the mean, were excluded. Trimmed averages and standard deviations were calculated for the remaining data. Rates were considered to be significantly different from the average if $R^{CSM/CSM}$ - $\delta R^{CSM/CSM}$ was larger than $<R^{CSM/CSM}> + \sigma$, or if $R^{CSM/CSM} + \delta R^{CSM/CSM}$ was smaller than $<R^{CSM/CSM}> - \sigma$ ($\Box R^{CSM/CSM}$ is the experimental error).

Pulse sequences

The pulse sequence used to measure multiple-quantum relaxation rates via direct detection of C' nuclei is shown in Fig. 1. This sequence is a modified version of a CON experiment[38] where C' coherence is converted via an INEPT-type building block into longitudinal two-spin order C'_zN_z. After a purging field gradient pulse, a coherence C'_xN_x is created by two simultaneous 90° pulses. This corresponds to a superposition of zero- and double-quantum terms:

$$2C'_xN_x = (C'_+N_- + C'_-N_+) + (C'_+N_+ + C'_-N_-)$$

The C'_xN_x coherence decays and is partly converted into C'_yN_y coherence during the mixing time τ_m, which can be expressed in terms of zero- and double-quantum relaxation rates R_{DQ} and R_{ZQ}:

$$\begin{aligned} 2C'_xN_x \rightarrow\ & C'_xN_x\,[\,\exp(-R_{ZQ}\tau_m)(C'_+N_- + C'_-N_+) + \exp(-R_{DQ}\tau_m)(C'_+N_+ + C'_-N_-)\,] \\ & + C'_yN_y\,[\,\exp(-R_{ZQ}\tau_m)(C'_+N_- + C'_-N_+) - \exp(-R_{DQ}\tau_m)(C'_+N_+ + C'_-N_-)\,] \end{aligned} \quad (2)$$

This conversion is driven by cross-correlated fluctuations. At the end of the mixing time

τ_m, the C'_xN_x and C'_yN_y terms are recorded in two separate experiments using 90° shifts of the phases ϕ_3 and ϕ_4. The ^{15}N magnetization is then allowed to evolve, prior to the transfer to ^{13}C coherence for detection. To obtain C' singlets, the effects of homonuclear $J(C'C^{\alpha})$ couplings are removed via an IPAP scheme during acquisition[38]. Each experiment was carried out twice, with different timings of the two final 180° C^{α} pulses and different phases of the last 90° C' pulses. The two subspectra give in-phase and anti-phase doublets, which are added and subtracted to separate the two multiplet components. These are then shifted to the centre of the original multiplet by $\pm$ ½$J(C'C^{\alpha})$ and summed together to obtain a singlet. The ^1H spins are decoupled throughout the entire sequence, whilst C^{α} 180° pulses are applied during the mixing time τ_m to suppress undesired scalar couplings and interference effects involving C^{α} spins. The relative signal intensities of the auto- and cross-correlated experiments:

$$<2C'_yN_y>/<2C'_xN_x> = \tanh(\Delta R\, \tau_m), \quad (3)$$

depend on the difference in rates $\Delta R = ½[R_{DQ} - R_{ZQ}]$ that corresponds to differential line-broadening.

Chemical shift modulation

Several studies[7,16,17,26,39] have shown that the difference in rates $\Delta R = ½[R_{DQ} - R_{ZQ}]$ can be decomposed into three terms:

$$\Delta R = R^{CSM/CSM} + R^{DD/DD} + R^{CSA/CSA} \quad (4)$$

The $R^{CSM/CSM}$ contribution accounts for fluctuations of isotropic chemical shifts (CS) that affect simultaneously the environments of C' and N nuclei of the same peptide plane. They are characterized by time-scales that typically fall in the μs to ms range[16].

These CS fluctuations can be caused by changes of dihedral ψ and ϕ backbone angles, as can be rationalized by *ab initio* calculations.[40] In the case of metal-binding proteins, they can also result from interactions with metal ions. The sign of the $R^{CSM/CSM}$ contribution depends on the signs of the changes of the C' and N chemical shifts upon conformational exchange. The modulations of the isotropic shifts are said to be *correlated* or *anti-correlated* if the $R^{CSM/CSM}$ rates are positive or negative.

The second term in Eq. (4) comprises contributions from various auto- and cross-correlated dipole-dipole effects.[41] As discussed elsewhere[8] the contributions to $R^{DD/DD}$ due to various dipolar auto-relaxation processes can safely be neglected, and and cross-relaxation processes are estimated from the 3D structure. The third term in Eq. (4) arises from cross-correlated CSA relaxation of C' and N nuclei due to concerted modulations of the chemical shifts that are caused by molecular motions. A study of cross-correlated relaxation in ubiquitin[42] showed that some of the principal components of the CSA tensors are, to a great extent, independent of the environment, while others are correlated with the isotropic shifts. These empirical correlations were found to be largely independent of the model for the peptide plane motions. Thus, for C' nuclei in ubiquitin, the σ_{yy} component of the CSA tensor, which is nearly parallel to the C=O bond, turns out to be roughly proportional to the isotropic chemical shift ($\sigma_{yy} = 3\sigma_{iso} - 334.9$ ppm) while the other two components are almost invariant ($\sigma_{xx} = 251.2$ ppm and $\sigma_{zz} = 83.6$ ppm). For amide ^{15}N nuclei, one finds that $\sigma_{yy} = 2\sigma_{iso} - 163.2$ ppm, $\sigma_{xx} = \sigma_{iso} + 105.5$ ppm and $\sigma_{zz} = 57.7$ ppm. Assuming that these empirical rules, originally derived for ubiquitin[42], are also applicable to the proteins under study, we obtain ~~a~~ rough estimates of the site-specific $R^{CSA/CSA}$ contributions and hence determine the "corrected" exchange rates $R^{CSM/CSM}$. This approach is similar to the analyses of cross-

correlated relaxation in Major Urinary Protein (MUP) in the presence or absence of a pheromone[16] and of the human centrin 2 in complex with the target peptide XPC.[8]

Acknowledgements We thank Marco Fragai (CERM, University of Florence), Andrea Battistoni and Marco Sette (University of Rome "Tor Vergata") for providing doubly labeled samples of MMP12 and Cu,Zn-SOD. This work was supported by the European Commission ("European Network of Research Infrastructures for providing Access and Technological Advancements in bio-NMR", EUNMR Contract # RII3-026145).

Supporting Information The $R^{CSM/CSM}$ rates for CaLaCb, Cu,ZnSOD and MMP12 are shown in Table S1.

References

1. Levitt, M., Sander, C., & Stern, P. (1985) Protein normal-mode dynamics: Trypsin inhibitor, crambin, ribonuclease and lysozyme. *J Mol Biol* 181:423-447.

2. Halioglu, T. & Bahar, I. (1999) Structure-based analysis of protein dynamics: comparison of theoretical results for hen lysozyme with X-ray diffraction and NMR relaxation data. *Proteins* 37:654-667.

3. Abergel, D. & Bodenhausen, G. (2005) Predicting internal protein dynamics from structures using coupled networks of hindered rotators. *J Chem Phys* 123:204901.

4. Bertini, I., Duma, L., Felli, I. C., Fey, M., Luchinat, C., Pierattelli, R., & Vasos, P. R. (2004) A heteronuclear direct detection NMR experiment for protein backbone assignment. *Angew. Chem. Int. Ed.* 43:2257-2259.

5. Bermel, W., Bertini, I., Felli, I. C., Piccioli, M., & Pierattelli, R. (2006) ^{13}C-detected *protonless* NMR spectroscopy of proteins in solution. *Progr. NMR Spectrosc.* 48:25-45.

6. Balayssac, S., Bertini, I., Luchinat, C., Parigi, G., & Piccioli, M. (2006) ^{13}C direct detected NMR increases the detectability of residual dipolar couplings. *J. Am. Chem. Soc.* 128:15042-15043.

7. Frueh, D., Tolman, J. R., Bodenhausen, G., & Zwahlen, C. (2001) Cross-correlated Chemical Shift Modulation: a signature of slow internal motions in proteins. *J. Am. Chem. Soc.* 123:4810.

8. Kateb, F., Abergel, D., Blouquit, Y., Duchambon, P., Craescu, C. T., & Bodenhausen, G. (2006) Slow backbone dynamics of the C-terminal fragment of human centrin 2 in complex with a target peptide probed by cross-correlated relaxation in multiple-quantum NMR spectroscopy. *Biochemistry* 45:15011-15019.

9. Kordel, J., Skelton, N. J., Akke, M., Palmer, A. G., III, & Chazin, W. J. (1992) Backbone dynamics of calcium-loaded calbindin D_{9k} studied by two-dimensional proton-detected ^{15}N NMR spectroscopy. *Biochemistry* 31:4856-4866.

10. Akke, M., Drakenberg, T., & Chazin, W. J. (1992) Three-dimensional solution structure of Ca^{2+}-loaded porcine calbindin D_{9K} determined by Nuclear Magnetic Resonance spectroscopy. *Biochemistry* 231:1011-1020.

11. Bertini, I., Donaire, A., Jiménez, B., Luchinat, C., Parigi, G., Piccioli, M., & Poggi, L. (2001) Paramagnetism-based versus classical constraints: an analysis of the solution structure of Ca Ln Calbindin D_{9k}. *J. Biomol. NMR* 21:85-98.

12. Bertini, I., Carrano, C. J., Luchinat, C., Piccioli, M., & Poggi, L. (2002) A ^{15}N NMR mobility study on the Di-calcium P43M calbindin D_{9K} and its mono La^{3+} substituted form. *Biochemistry* 41:5104-5111.

13. Allegrozzi, M., Bertini, I., Janik, M. B. L., Lee, Y.-M., Liu, G., & Luchinat, C. (2000) Lanthanide induced pseudocontact shifts for solution structure refinements of macromolecules in shells up to 40 Å from the metal ion. *J. Am. Chem. Soc.* 122:4154-4161.

14. Bertini, I., Lee, Y.-M., Luchinat, C., Piccioli, M., & Poggi, L. (2001) Locating the metal ion in calcium-binding proteins by using cerium(III) as a probe. *ChemBioChem* 2:550-558.

15. Bertini, I., Jiménez, B., Piccioli, M., & Poggi, L. (2005) Asymmetry in ^{13}C-^{13}C COSY spectra identifies geometry in paramagnetic proteins. *J. Am. Chem. Soc.* 127:12216-12217.

16. Perazzolo, C., Wist, J., Loth, K., Poggi, L., Homans, S. W., & Bodenhausen, G. (2005) Effects of protein-pheromone complexation on correlated chemical shift modulations. *J. Biomol. NMR* 33:233-242.

17. Wist, J., Perazzolo, C., & Bodenhausen, G. (2005) Slow motions in nondeuterated proteins: concerted chemical shift modulations of backbone nuclei. *Appl. Magn. Reson.* 29:251-259.

18. Gabbianelli R., D'Orazio, M., Pacello, F., O'Neill, P., Nicolini L., Rotilio, G., & Battistoni, A. (2004) Distinctive functional features in prokaryotic and eukaryotic Cu,Zn-superoxide dismutases. *Biol. Chem.* 385:749-754.

19.Mori, M., Jiménez, B., Piccioli, M., Battistoni, A., & Sette, M. (2008) The solution structure of the monomeric Copper, Zinc Superoxide Dismutase from *Salmonella enterica*: structural insights to understand the evolution toward the dimeric structure. *Biochemistry* 47:12954-12963.

20. Jiménez, B., Mori, M., Battistoni, A., Sette, M., & Piccioli, M. (2007) NMR assignment of reduced form of copper, zinc superoxide dismutase from *Salmonella enterica*. *Biomol. NMR Assign.* 1:65-68.

21. Pesce, A., Capasso, C., Battistoni, A., Folcarelli, S., Rotilio, G., Desideri, A., & Bolognesi, M. (1997) Unique structural features of the Monomeric Cu,Zn-Superoxide Dismutase from Escherichia coli, revealed by X-ray crystallography. *J. Mol. Biol.* 274:408-420.

22. Johnson, L. L., Dyer, R., & Hupe, D. J. (1998) Matrix metalloproteinases. *Curr. Opin. Chem. Biol.* 2:466-471.

23. Bode, W., Fernandez-Catalan, C., Tschesche, H., Grams, F., Nagase, H., & Maskos, K. (1999) Structural properties of matrix metalloproteinases. *Cell Mol. Life Sci.* 55:639-652.

24. Bertini, I., Calderone, V., Cosenza, M., Fragai, M., Lee, Y.-M., Luchinat, C., Mangani, S., Terni, B., & Turano, P. (2005) Conformational variability of MMPs: beyond a single 3D structure. *Proc. Natl. Acad. Sci. USA* 102:5334-5339

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

26. Frueh, D. (2002) Internal motions in proteins and interference effects in nuclear magnetic resonance. *Prog. NMR Spectrosc.* 41:305-324.

27. Wishart, D. S. & Case, D. A. (2001) Use of chemical shifts in macromolecular structure determination. *Methods Enzymol* 338:3-34.

28. Oldfield, E. (2002) Chemical shifts in amino acids, peptides and proteins: from quantum chemistry to drug design. *Annu. Rev. Phys. Chem.* 53:349-378

29. Xu, X. P. & Case, D. A. (2002) Probing multiple effects on 15N, 13C alpha, 13C beta, and 13C' chemical shifts in peptides using density functional theory. *Biopolymers* 65:408-423.

30. Emberly, E. G., Mukhopadhyay, R., Tang, C., & Wingreen, N. S. (2004) Flexibility of beta-sheets: principal component analysis of database protein structures. *Proteins* 55:91-98.

31. Emberly, E. G., Mukhopadhyay, R., Wingreen, N. S., & Tang, C. (2003) Flexibility of alpha-helices: results of a statistical analysis of database protein structures. *J Mol Biol* 327:229-237.

32. Le, H. & Oldfield, E. (1996) *Ab Initio* studies of amide-^{15}N Chemical Shifts in dipeptides: applications to protein NMR spectroscopy. *J Phys Chem* 100:16423-16428.

33. Le, H. & Oldfield, E. (1994) Correlation between ^{15}N NMR chemical shifts in proteins and secondary structure. *J Biomol NMR* 4:341-348.

34. Wishart, D. S. & Case, D. A. (2001) Use of chemical shifts in macromolecular structure determination. *Methods Enzymol* 338:3-34.

35. Wishart, D. S. & Sykes, B. D. (1994) Chemical shifts as a tool for structure determination. *Methods Enzymol* 239:363-392.

36. Chazin, W. J., Kördel, J., Drakenberg, T., Thulin, E., Brodin, P., Grundstrom, T., & Forsén, S. (1989) Proline isomerism leads to multiple folded conformations of Calbindin D_{9k}: direct evidence from two-dimensional ^{1}H NMR

spectroscopy. *Proc. Natl. Acad. Sci. USA* 86:2195-2198.

37. Banci, L., Bertini, I., Ciulli, A., Fragai, M., Luchinat, C., & Terni, B. (2003) Expression and high yield production of Matrix Metalloproteinase 12 and of an active mutant with increased solubility. *J. Mol. Catal. A: Chemical* 204-205:401-408.

38. Bermel, W., Bertini, I., Duma, L., Emsley, L., Felli, I. C., Pierattelli, R., & Vasos, P. R. (2005) Complete assignment of heteronuclear protein resonances by protonless NMR spectroscopy. *Angew. Chem. Int. Ed.* 44:3089-3092.

39. Dittmer, J. & Bodenhausen, G. (2004) Evidence for slow motion in proteins by multiple refocusing of heteronuclear nitrogen/proton multiple quantum coherences in NMR. *J Am Chem Soc.* 126:1314-1315.

40. Arnold, W. D. & Oldfield, E. (2000) The chemical nature of hydrogen bonding in proteins via NMR: *J*-Couplings, Chemical Shifts, and AIM theory. *J Am Chem Soc.* 122:12835-12841.

41. Konrat, R. & Sterk, H. (1993) Cross-correlation effects in the transverse relaxation of multiple-quantum transitions of heteronuclear spin systems. *Chem Phys Lett.* 203:75.

42. Loth, K., Pelupessy, P., & Bodenhausen, G. (2005) Chemical shift anisotropy tensors of carbonyl, nitrogen, and amide proton nuclei in proteins through cross-correlated relaxation in NMR spectroscopy. *J Am Chem Soc.* 127:6062-6068.

Figure Legends

Figure 1. Pulse sequence designed to measure relaxation rates of $^{13}C'$-^{15}N multiple-quantum coherences. All pulses were applied along the x axes, unless otherwise indicated. Black and white pulses applied to the ^{15}N spins indicate non-selective 90° and 180° pulses, respectively. All ^{13}C pulses were band selective; Q5 (or time-reversed Q5) shaped 90° ^{13}C pulses and Q3 shaped 180° ^{13}C pulses were used with durations of 256 and 220 μs. For $^{13}C^{\alpha}$ inversion pulses, Q3 shaped pulses were used with durations of 220 μs. The ^{13}C, ^{15}N, and ^{1}H carrier frequencies were centred at 173, 116 and 3.5 ppm, respectively. ^{1}H decoupling was performed with a 2.9 kHz WALTZ-16 sequence. ^{15}N decoupling during acquisition was performed using a 2.9 kHz GARP-4 sequence. To suppress $J(C'C_{\alpha})$ couplings during acquisition, the in-phase (IP) and anti-phase (AP) components were acquired and stored separately. Pulses depicted in grey were used to acquire anti-phase components, whilst pulses represented by hatched shapes were used to obtain in-phase components[12]. The delays Δ, δ, ε, were set to 12.5, 1.75 and 4.5 ms, respectively. The mixing times were τ_m = 20 and 40 ms. The durations of all sine-shaped gradients were 1 ms and the peak amplitudes were g_{1z} = 57, g_{2z} = 37, g_{3z} = -29, g_{4z} = -17, g_{5z} = 19, and g_{6z}= 41 Gcm^{-1}. The phases for the auto-relaxation experiment were: ϕ_1 = y, -y; ϕ_2 = y, y, -y, -y; ϕ_3 = 16 y, 16 -y; ϕ_4 = 4y, 4 -y ;ϕ_5 = 8 x, 8 -x; ϕ_6 = 32 x, 32 -x; ϕ_7 = x for in phase acquisition, ϕ_7 = -y for antiphase acquisition; ϕ_{rec} = (x, -x, -x, x, -x, x, x, -x), 2(-x, x, x, -x, x, -x, -x, x), (x, -x, -x, x, -x, x, x, -x), (-x, x, x, -x, x, -x, -x, x), 2(x, -x, -x, x, -x, x, x, -x), (-x, x, x, -x, x, -x, -x, x). The phases for the cross-relaxation experiment were: ϕ_1 = y, -y; ϕ_2 = y, y, -y, -y; ϕ_3 = 16 x, 16 -x; ϕ_4 = 4 x, 4 -x; ϕ_5 = 8 x, 8 -x; ϕ_6 = 32 x, 32 -x; ϕ_7 = x for in phase acquisition; ϕ_7 = -y for antiphase

acquisition; ϕ_{rec}= (x, -x, -x, x, -x, x, x, -x), 2(-x, x, x, -x, x, -x, -x, x), (x, -x, -x, x, -x, x, x, -x), (-x, x, x, -x, x, -x, -x, x), 2(x, -x, -x, x, -x, x, x, -x), (-x, x, x, -x, x, -x, -x, x).

Figure 2. The difference in the relaxation rates $\Delta R = ½[R_{DQ} - R_{ZQ}]$ of C'N double- and zero-quantum coherences measured either via the novel 'protonless' ^{13}C-detected experiments (^{13}C ΔR, shown vertically) or via established ^{1}H-detected experiments (^{1}H ΔR, shown horizontally) for both diamagnetic CaLaCb and paramagnetic CaCeCb. The error bars represent experimental uncertainties in both experiments.

Figure 3. Comparison of the novel 'protonless' ^{13}C-detected, and established ^{1}H-detected experiments for paramagnetic CaCeCb at 700 MHz, 298 K, with τ_m = 40 ms. (A) Row corresponding to i = 44 (residues S44 and T45) of a 'protonless' ^{13}C-detected CON experiment to measure relaxation of the C'_iN_{i+1} coherence. (B) Matching row (i + 1= 45) of the corresponding ^{1}H-detected HNCO experiment. The ranges of chemical shifts correspond to 12 ppm for ^{13}C and 3 ppm for ^{1}H (both spectral widths are 2.1 kHz.) (C) Signal corresponding to i = 55 obtained by the a 'protonless' ^{13}C-detected CON version of the experiment, which could not be observed with the ^{1}H-detected experiment.

Figure 4. $R^{CSM/CSM}$ rates obtained for the proteins (A) CaLaCb, (B) Cu,ZnSOD and (C) MMP12. The background colour indicates the secondary structure: green for α-helices and blue for β-sheets. In CaLaCb, local deviations of $R^{CSM/CSM}$ from the average are observed in the two metal binding loops. Residues that coordinate the metal ion in loops I (residues 17 and 22) and II (residues 56, 58 and 65) are highlighted in red and blue,

depending on whether metal binding occurs through the backbone carbonyl oxygen or through side-chain carboxylate groups. For CaLaCb, dashed lines indicate the standard deviations with respect to the average $\langle R^{CSM/CSM} \rangle$ of the entire protein. For Cu,ZnSOD and MMP12 (panels B and C), the dashed lines denote *local* averages and standard deviations for individual secondary structure elements (α-helices and β-sheets). The numbers of the amino acids refer to the carbonyl atoms, i.e., i = 54 correspond to double- and zero-quantum coherences involving C'(i = 54) and N(i = 55).

Figure 5. Overview of all $R^{CSM/CSM}$ rates determined in CaLaCb, Cu,Zn-SOD and MMP12, plotted as a function of the backbone dihedral angle ψ. The clusters to the left and right correspond to α helices and β sheets.

Figure 1.

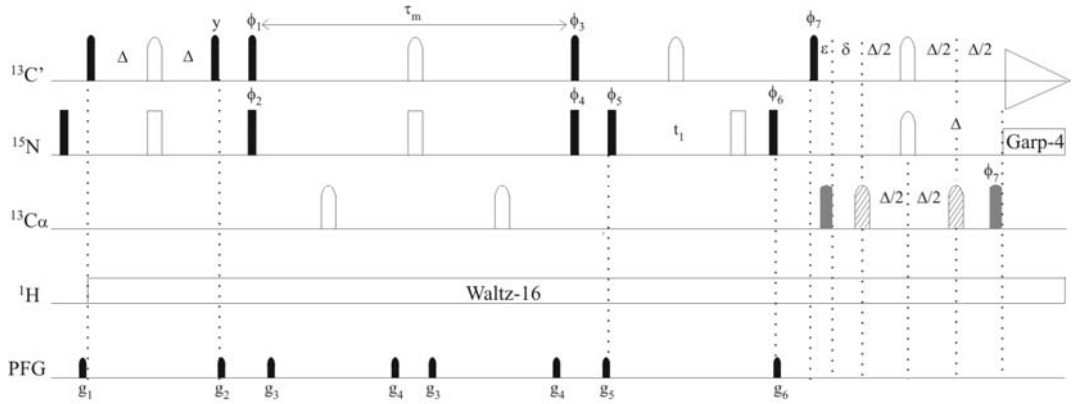
^{13}C'
^{15}N
^{13}Cα
^{1}H
PFG
Δ
Δ
y
ϕ_1
τ_m
ϕ_3
ϕ_7
ε
δ
Δ/2
Δ/2
Δ/2
ϕ_2
ϕ_4
ϕ_5
ϕ_6
t_1
Δ
Garp-4
ϕ_7
Δ/2
Δ/2
Waltz-16
g_1
g_2
g_3
g_4
g_3
g_4
g_5
g_6

Figure 2.

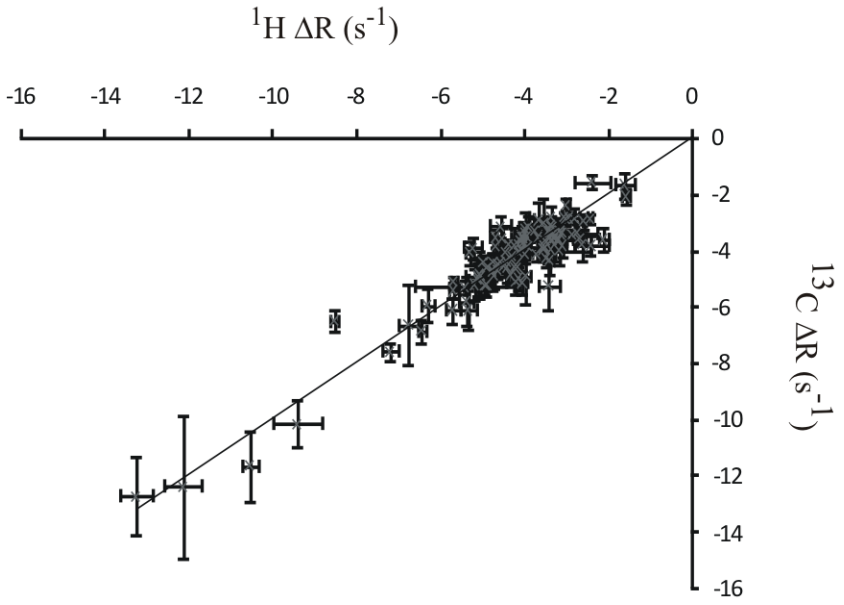

Figure 3.

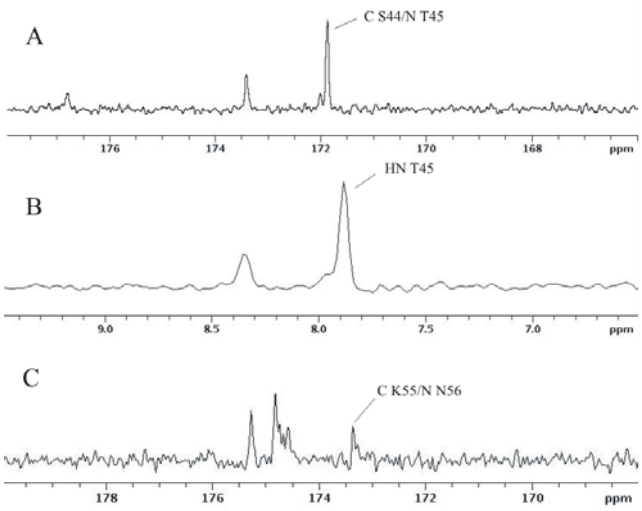
A
C S44/N T45
176
174
172
170
168
ppm
B
HN T45
9.0
8.5
8.0
7.5
7.0
ppm
C
C K55/N N56
178
176
174
172
170
ppm

Figure 4.

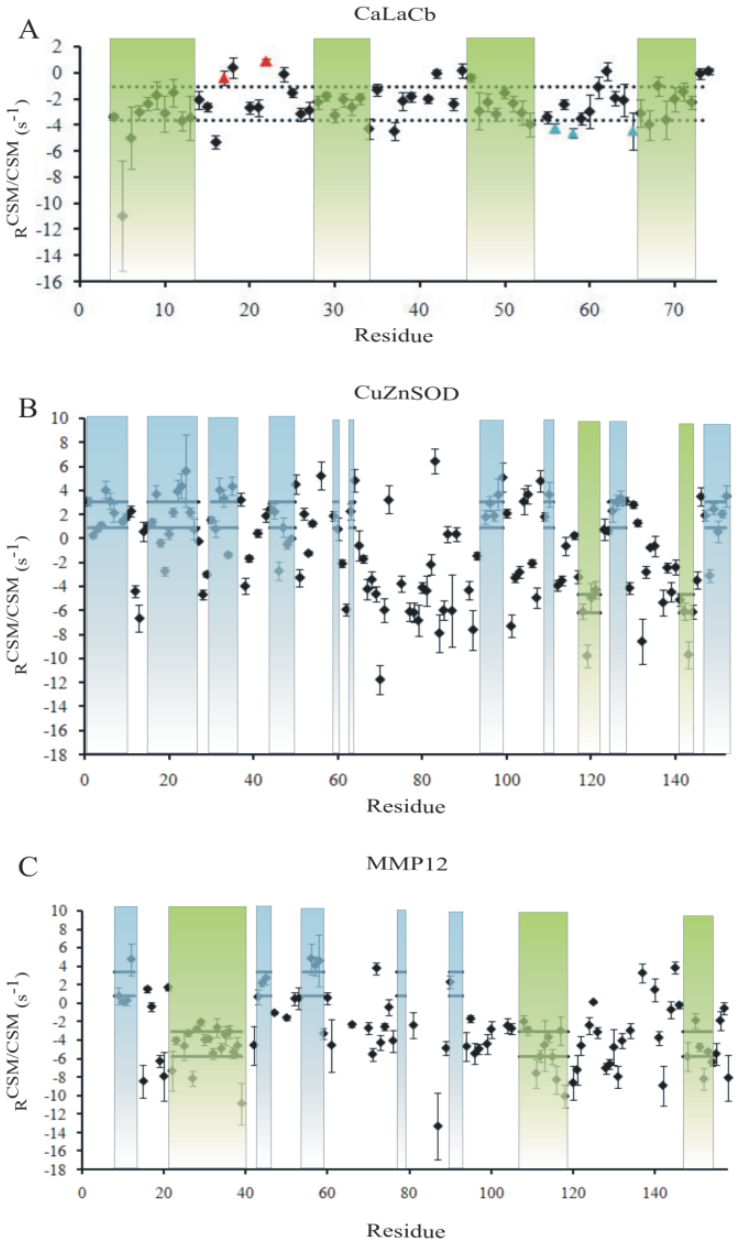
A
CaLaCb
$R^{CSM/CSM}$ (s^{-1})
Residue
B
CuZnSOD
$R^{CSM/CSM}$ (s^{-1})
Residue
C
MMP12
$R^{CSM/CSM}$ (s^{-1})
Residue

Figure 5.

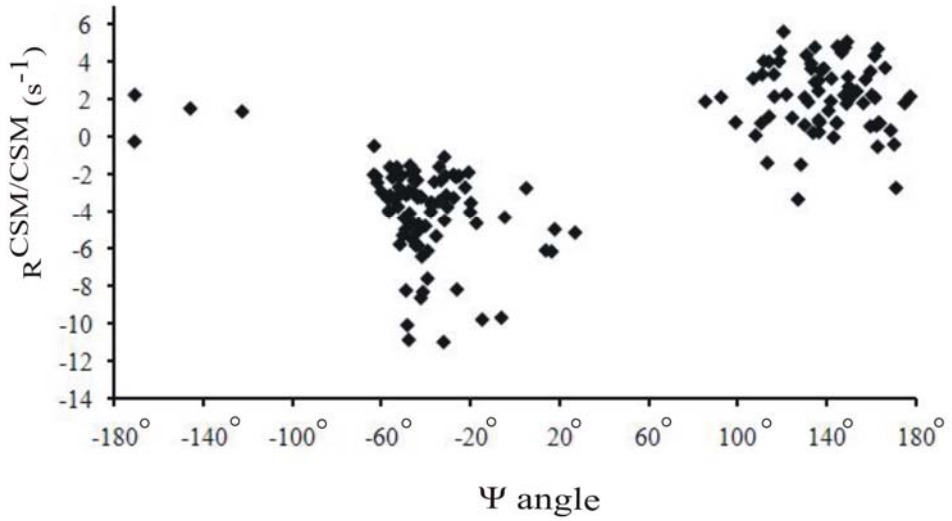
$R^{CSM/CSM}$ (s^{-1})
6
4
2
0
-2
-4
-6
-8
-10
-12
-14
-180°
-140°
-100°
-60°
-20°
20°
60°
100°
140°
180°
Ψ angle

Table S1

CaLaCb			CuZnSOD			MMP12		
Residues		**$R^{CSM/CSM}$ (sec-1)**	**Residues**		**$R^{CSM/CSM}$ (sec-1)**	**Residues**		**$R^{CSM/CSM}$ (sec-1)**
4	GLU	-3.40	1	ALA	3.09	9	ILE	0.76
5	GLU	-10.97	2	SER	0.26	10	THR	0.11
6	LEU	-5.01	3	GLU	0.81	11	TYR	0.28
7	LYS	-3.04	4	LYS	1.12	12	ARG	4.73
8	GLY	-2.39	5	VAL	4.04	15	ASN	-8.44
9	ILE	-1.75	6	GLY	3.15	16	TYR	1.50
10	PHE	-3.11	7	MET	2.14	17	THR	-0.39
11	GLU	-1.53	9	LEU	1.43	19	ASP	-6.27
12	LYS	-3.71	10	VAL	1.84	20	MET	-7.91
13	TYR	-3.46	11	THR	2.29	21	ASN	1.68
14	ALA	-2.09	12	ALA	-4.39	22	ARG	-7.33
15	ALA	-2.62	13	GLN	-6.64	23	GLU	-4.04
16	LYS	-5.33	14	GLY	0.59	25	VAL	-4.64
17	GLU	-0.37	15	VAL	1.05	26	ASP	-3.26
18	GLY	0.37	16	GLY	1.38	27	TYR	-8.15
20	PRO	-2.69	17	GLN	3.71	28	ALA	-2.82
21	ASN	-2.67	18	SER	-0.37	29	ILE	-2.06
22	GLN	0.89	19	ILE	-2.71	30	ARG	-3.96
24	SER	-0.13	20	GLY	0.38	31	LYS	-3.87
25	LYS	-1.55	21	THR	2.19	32	ALA	-5.42
26	GLU	-3.17	22	VAL	3.94	33	PHE	-2.64

27	GLU	-2.89	23	VAL	4.39	34	GLN	-4.93
28	LEU	-2.25	24	ILE	5.64	35	VAL	-3.45
29	LYS	-1.80	25	ASP	2.18	36	TRP	-3.07
30	LEU	-3.26	26	GLU	0.76	37	SER	-5.29
31	LEU	-2.05	27	THR	-0.22	38	ASN	-4.59
32	LEU	-2.64	28	GLU	-4.66	39	VAL	-10.85
33	GLN	-1.96	29	GLY	-2.98	42	LEU	-4.55
34	THR	-4.30	30	GLY	1.55	43	LYS	0.67
35	GLU	-1.34	31	LEU	0.80	44	PHE	2.15
37	PRO	-4.49	32	LYS	4.05	45	SER	2.71
38	SER	-2.18	33	PHE	3.36	47	ILE	-1.06
39	LEU	-1.84	34	THR	-1.35	50	GLY	-1.58
41	LYS	-2.02	35	PRO	4.35	52	ALA	0.48
42	GLY	-0.07	37	LEU	3.23	53	ASP	0.53
44	SER	-2.41	38	LYS	-3.96	56	VAL	4.81
45	THR	0.15	39	ALA	-1.65	57	VAL	4.06
46	LEU	-0.42	41	PRO	0.44	58	PHE	4.57
47	ASP	-2.95	43	GLY	1.91	59	ALA	-3.30
48	GLU	-2.27	44	GLU	2.49	60	ARG	0.56
49	LEU	-3.18	45	HIS	2.25	61	GLY	-4.56
50	PHE	-1.55	46	GLY	-2.69	66	ASP	-2.32
51	GLU	-2.35	47	PHE	0.93	70	ASP	-2.71
52	GLU	-3.07	48	HIS	-0.49	71	GLY	-5.53
53	LEU	-3.96	49	ILE	0.01	72	LYS	3.76
55	LYS	-3.39	50	HIS	4.53	73	GLY	-4.28
56	ASN	-4.26	51	ALA	-3.25	74	GLY	-2.56
57	GLY	-2.44	52	ASN	2.05	75	ILE	-0.45
58	ASP	-4.61	53	GLY	-1.22	76	LEU	-4.07

59	GLY	-3.53	54	SER	1.25	81	GLY	-2.40
60	GLU	-2.99	56	GLN	5.24	87	GLY	-13.30
61	VAL	-1.11	59	ILE	1.90	89	ASP	-4.87
62	SER	0.10	60	LYS	0.79	90	ALA	2.26
63	PHE	-1.96	61	ASP	-2.08	94	GLU	-4.69
64	GLU	-2.12	62	GLY	-5.90	95	ASP	-1.69
65	GLU	-4.44	63	GLN	2.27	96	GLU	-5.46
66	PHE	-3.12	64	ALA	4.86	97	PHE	-4.95
67	GLN	-3.98	65	VAL	-0.58	99	THR	-4.42
68	VAL	-1.01	66	ALA	-1.68	100	THR	-2.83
69	LEU	-3.62	67	ALA	-4.20	104	GLY	-2.44
70	VAL	-2.02	68	GLU	-3.40	105	THR	-2.75
71	LYS	-1.47	69	ALA	-4.62	108	PHE	-2.00
72	LYS	-2.27	70	ALA	-11.75	109	LEU	-2.92
73	ILE	-0.08	71	GLY	-5.95	111	ALA	-7.57
74	SER	0.11	72	GLY	3.23	112	VAL	-5.74
			75	ASP	-3.75	113	HIS	-4.55
			77	GLN	-6.08	114	GLU	-3.68
			78	ASN	-6.17	115	ILE	-5.80
			79	THR	-6.82	116	GLY	-8.28
			80	GLY	-4.07	117	HIS	-2.98
			81	LYS	-4.35	118	SER	-10.06
			82	HIS	-2.16	120	GLY	-8.61
			83	GLU	6.46	121	LEU	-7.21
			84	GLY	-7.88	122	GLY	-4.60
			85	PRO	-5.97	124	SER	-2.41
			86	GLU	0.38	125	SER	0.10
			87	GLY	-6.01	126	ASP	-3.14

88	GLN	0.38	128	LYS	-7.03
91	LEU	-4.29	129	ALA	-6.59
92	GLY	-7.59	130	VAL	-4.77
93	ASP	-1.44	131	MET	-7.95
95	PRO	1.80	132	PHE	-4.08
96	VAL	2.95	134	THR	-2.97
97	LEU	1.92	137	TYR	3.25
98	VAL	3.67	140	ILE	1.45
99	VAL	5.10	141	ASN	-3.74
100	ASN	2.11	142	THR	-8.91
101	ASN	-7.29	144	ARG	-0.72
102	ASP	-3.26	145	LEU	3.81
103	GLY	-2.79	146	SER	-0.22
104	ILE	3.09	148	ASP	-5.79
105	ALA	3.69	150	ILE	-1.88
106	THR	-2.05	151	ARG	-4.74
107	GLU	-4.93	152	GLY	-8.20
108	PRO	4.82	153	ILE	-5.25
109	VAL	1.83	154	GLN	-6.39
110	THR	3.68	155	SER	-5.49
112	PRO	-3.90	156	LEU	-1.89
113	ARG	-3.49	157	TYR	-0.57
114	LEU	-0.62	158	GLY	-8.08
116	SER	0.26			
117	LEU	-3.21			
118	ASP	-6.08			
119	GLU	-9.77			
120	VAL	-4.93			

121	LYS	-4.27
123	LYS	0.76
124	ALA	0.64
125	LEU	2.31
126	MET	3.05
127	ILE	3.36
128	HIS	3.14
129	VAL	-4.11
130	GLY	2.83
131	GLY	1.31
132	ASP	-8.57
133	ASN	-2.78
134	MET	-0.74
135	SER	-0.61
137	GLN	-5.33
138	PRO	-2.42
139	LYS	-4.46
140	PRO	-2.38
141	LEU	-5.10
142	GLY	-6.06
143	GLY	-9.67
144	GLY	-6.12
145	GLY	-3.47
146	THR	3.51
147	ARG	1.94
148	TYR	-3.08
149	ALA	2.47
150	CYS	0.59

	151	GLY	2.07
	152	VAL	3.55