Analysis of ¹³Cα and ¹³Cβ chemical shifts of cysteine and cystine residues in proteins: a quantum chemical approach

Osvaldo A. Martin,
Instituto de Matemática Aplicada San Luis, Universidad Nacional de San Luis, CONICET, Ejército de Los Andes 950, 5700 San Luis, Argentina

Myriam E. Villegas,
Instituto de Matemática Aplicada San Luis, Universidad Nacional de San Luis, CONICET, Ejército de Los Andes 950, 5700 San Luis, Argentina

Jorge A. Vila, and
Instituto de Matemática Aplicada San Luis, Universidad Nacional de San Luis, CONICET, Ejército de Los Andes 950, 5700 San Luis, Argentina

Baker Laboratory of Chemistry and Chemical Biology, Cornell University, Ithaca, NY 14853-1301, USA

Harold A. Scheraga
Baker Laboratory of Chemistry and Chemical Biology, Cornell University, Ithaca, NY 14853-1301, USA

Harold A. Scheraga: has5@cornell.edu

Abstract

Cysteines possess a unique property among the 20 naturally occurring amino acids: it can be present in proteins in either the reduced or oxidized form, and can regulate the activity of some proteins. Consequently, to augment our previous treatment of the other types of residues, the ¹³Cα and ¹³Cβ chemical shifts of 837 cysteines in disulfide-bonded cystine from a set of seven non-redundant proteins, determined by X-ray crystallography and NMR spectroscopy, were computed at the DFT level of theory. Our results indicate that the errors between observed and computed ¹³Cα chemical shifts of such oxidized cysteines can be attributed to several effects such as: (a) the quality of the NMR-determined models, as evaluated by the conformational-average (*ca*) rmsd value; (b) the existence of high B-factor or crystal-packing effects for the X-ray-determined structures; (c) the dynamics of the disulfide bonds in solution; and (d) the differences in the experimental conditions under which the observed ¹³Cα chemical shifts and the protein models were determined by either X-ray crystallography or NMR-spectroscopy. These quantum-chemical-based calculations indicate the existence of two, almost non-overlapped, basins for the oxidized and reduced –SH ¹³Cβ, but not for the ¹³Cα, chemical shifts, in good agreement with the observation of 375 ¹³Cα and 337 ¹³Cβ resonances from 132 proteins by Sharma and Rajarathnam (2000). Overall, our results indicate that explicit consideration of the disulfide bonds is a necessary condition for an accurate prediction of ¹³Cα and ¹³Cβ chemical shifts of cysteines in cystines.

Correspondence to: Harold A. Scheraga, has5@cornell.edu.

Keywords

Introduction

Cystine (with disulfide bonds formed) and cysteine residues are important for protein structure and function. In their reduced form, cysteines participate in the active site of different enzymes and, together with histidine, are the two most common residues involved in the coordination of zinc ions in Zn-finger motifs (which are the most commonly observed structural motifs in transcription factors; Kornhaber et al. (2006). In the last few years, it has been recognized that disulfide bonds are not only inert structural elements. But, on the contrary, pairs of cysteines play an active role in the catalytic cycle of enzymes such as thioredoxin, switching between the reduced and oxidized forms. The activity of some proteins is regulated by the redox state of the cysteines or by their glutathionylation and nitrosylation (Wouters et al. 2007 and references therein).

In a seminal work, de Dios et al. (1993) showed that chemical shifts of proteins can be computed accurately by quantum chemical approaches. The motivation to compute $^{13}C^{\alpha}$ chemical shifts arises from the fact that they are exquisitely sensitive to, and depend mainly on, the backbone torsional angles (ϕ, ψ; Spera and Bax 1991), although the influence of the side-chain torsional angles, χ's, cannot be disregarded (Havlin et al. 1997; Pearson et al. 1997; Xu and Case 2001; Sun et al. 2002; Villegas et al. 2007; Vila et al. 2009). This property enables us to treat each residue **X** of a protein as a terminally-blocked tripeptide with the sequence Ac-G**X**G-NMe, with **X** in the conformation of the experimental protein structure and, hence, permitting the parallelization of the quantum-mechanical calculations. Following this procedure, $^{13}C^{\alpha}$ chemical-shift computations are feasible for proteins of any size and topology, e.g., by using the recently-introduced *Che*Shift server (Vila et al. 2009).

However, the above protocol to compute $^{13}C^{\alpha}$ chemical shifts for any naturally occurring amino acid in proteins breaks down for cystines because: (a) cystines cannot be treated as single tripeptides; (b) the cysteines involved in disulfide bonds usually are separated by at least four consecutive residues in the sequence, making quantum-chemical calculations impossible with existing computational resources, if all the intervening residues between the two bonded cysteines are taken into account; and (c) there is no evidence that the $^{13}C^{\alpha}$ chemical shifts for the cysteines of disulfide bonds can be obtained straightforwardly from those for reduced cysteine. For these reasons, previous calculations of $^{13}C^{\alpha}$ chemical shifts for cysteines, e.g., by using the *Che*Shift server, were *always* carried out by assuming that the cysteine residues were in their reduced form.

In order to compute the $^{13}C^{\alpha}$ chemical shifts of cysteines in disulfide bonds, in this work we present an extension of an existing protocol (Vila and Scheraga 2009) that will enable us to compute the $^{13}C^{\alpha}$ chemical shifts for cysteines in disulfide bonds accurately. In order to test this new methodology, seven protein models, rich in disulfide bonds, obtained by both NMR spectroscopy and X-ray crystallography, are considered. Each of these seven proteins was subjected to a $^{13}C^{\alpha}$-based chemical-shift validation analysis, i.e., by using the *ca*-rmsd (Vila et al. 2007; Vila and Scheraga 2009) as a scoring function. This analysis enables us to shed light on the origin of errors between observed and computed $^{13}C^{\alpha}$ chemical-shift values for cysteines in disulfide bonds as well as to detect the existence of local flaws in the amino acid sequence. Finally, a brief analysis of the dependence of $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ chemical shifts on the redox state of the cysteines was also carried out. This analysis enables us to determine how

well the DFT-based computations of the $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ chemical shifts agree with existing statistical-based analyses of observed $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ chemical-shift distributions as functions of the oxidation state of cysteines in proteins (Sharma and Rajarathnam 2000).

Materials and methods

Experimental set of structures

A set of seven proteins was considered in this work (see Table 1). Additional information regarding the set of structures used and how they were selected can be found in the Supplementary Material.

Method to compute $^{13}C^{\alpha}$ chemical shifts

The computations of the $^{13}C^{\alpha}$ chemical shifts involve a series of approximations: (a) for each cysteine residue **C** *not* involved in a disulfide bond, the computation of the $^{13}C^{\alpha}$ shielding was carried out on a terminally-blocked tripeptide with the sequence Ac-GCG-NMe, with **C** in the conformation of the regularized experimental protein structure, and the sulfhydryl group was protonated. Then, the $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ chemical shift for each amino acid residue **C** was computed at the OB98/6-311+G(2d,p) level of theory (Vila and Scheraga 2009), while the remaining residues in the tripeptide were treated at the OB98/3-21G level of theory, i.e., by using the *locally-dense* approach (Chesnut and Moore 1989; all the computed $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ shielding

values $\left(\sigma_{subst,th}^{n}, \text{with } n=\alpha \text{ or } \beta \right)$ were calculated by using the gauge-invariant atomic orbital (GIAO) method at the DFT level of theory as implemented in the GAUSSIAN 03 suite of programs (Frisch et al. 2004); (b) each of the cysteine residues $\mathbf{C}_i$ and $\mathbf{C}_j$ (with i and j denoting the position in the sequence) *forming* a disulfide bond were, first, treated as a terminally-blocked tripeptide, namely, Ac-G$\mathbf{C}_i$G-NMe and Ac-G$\mathbf{C}_j$G-NMe, respectively, with $\mathbf{C}_i$ and $\mathbf{C}_j$ *and* their disulfide group in the conformation of the regularized experimental protein structure, and protonated; secondly, the cysteines in cystine were treated as if in a hexapeptide in the computation, as shown in Fig. 1.

Further information regarding the regularized geometry adopted for the calculations and details of the method can be found in the Supplementary Material.

Computation of the conformationally-averaged rmsd (*ca*-rmsd)

in several previous papers (Vila et al. 2007, 2008, 2009; Vila and Scheraga 2008, 2009) and, hence, we reproduce here, for the reader's convenience only, the main definitions. For further details, see Supplementary Material.

Under the assumptions of fast conformational averaging, the computation of the *ca*-rmsd for a protein containing N amino acids residues, is given by (Vila et al. 2007):

$$ca-\mathrm{rmsd}^{n}=\left[(1/N) \sum_{\mu=1}^{N} \left(^{13}C_{observed,\mu}^{n} - <^{13}C_{computed,\mu}^{n}> \right)^{2} \right]^{1/2} \text{ with } n=\alpha \text{ or } \beta, \text{ and } <^{13}C_{computed}^{n}>_{\mu}=(1/\Omega) \sum_{i=1}^{\Omega} {}^{13}C_{\mu,i}^{n}$$

, for a given amino acid residue μ and with Ω the total number of protein conformations. Evidently, if $\Omega = 1$; *ca*-rmsd$^n \equiv$ rmsd^n; as for any single structure. In addition, for each amino acid μ, we define an error function $\Delta_{\mu}^{n}=(^{13}C_{observed,\mu}^{n} - {}^{13}C_{computed,\mu}^{n})$, with $n = \alpha$ *or* β.

A normalized rmsd for comparing different protein structures

In the absence of a *gold-standard*, it is common practice in the field of protein structure determination to compare NMR-derived conformations against a single X-ray derived structure. However, the corresponding X-ray structure does not always exist and, more important, even if it exists, a single X-ray structure may, or may not, be a better representation

of the observed $^{13}C^\alpha$ chemical shifts in solution than an NMR- or X-ray-determined ensemble of conformations (Vila and Scheraga 2009; Arnautova et al. 2009). Nevertheless, attempts to adopt the rmsd, between observed and predicted $^{13}C^\alpha$ chemical shifts for a protein structure solved at high accuracy, as a quality-model against which to compare other NMR-derived structures have a drawback. Among others, it is well known that the rmsd parameter is a reliable indicator of the global property of protein structures containing the same, or similar, number of residues (Maiorov and Crippen 1995; Betancourt and Skolnick 2001; Carugo and Pongor 2001). In other words, the rsmd is affected by the conformation similarity *and* the overall sizes of the proteins being compared (Maiorov and Crippen 1995). In the Supplementary Material section, we provide a discussion of this problem by analyzing the rmsd's of a set of 24 proteins, solved by NMR spectroscopy, with a broad number of residues (*N*) ranging from 48 to 370. A solution to this important problem lies beyond the goal of this manuscript and, hence, we adopted the expression proposed by Carugo and Pongor (2001), as a normalized size-independent rmsd, viz.,

$$\text{rmsd}_L = \frac{\text{rmsd}}{1+\ln\sqrt{N/L}}$$

(1)

where *N* is the number of residues in the sequence of any given protein, *L* is the number of residues in the protein chosen as a reference, and rmsd_L is the normalized, size-independent rmsd value that would be measured if the given structure under consideration contains *L* residues. It is worth noting that Eq. 1 breaks down for *N* lower than ~14 residues for *L* ~100, because the rmsd becomes negative. This is not a problem because we are dealing with proteins, not oligopeptides.

For the purpose of this work, we chose *L* = 76, as the reference residue number, i.e., with *L* representing the size of the ubiquitin protein, a highly-accurately-solved NMR protein structure, e.g., 1D3Z (Cornilescu et al. 1998), with a *ca*-rmsd of 2.20 ppm (Vila and Scheraga 2009). Use of Eq. 1 to compare the quality of NMR-derived ensembles of structures with different sizes, in terms of $^{13}C^\alpha$ chemical shifts, can be carried out by substituting *ca*-rmsd for rmsd and *ca*-rmsd$_L$ for rmsd$_L$. For this purpose, it is useful to define a *ca*-rmsd$_{76}$ cutoff value beyond which a need for further refinement of any given protein is necessary. Consequently, a *ca*-rmsd$_{76}$ = 2.6 ppm as a cutoff value was adopted (see Supplementary Material for details leading to this selection).

Results and discussion

Analysis of the NMR and X-ray conformations

A comparative analysis, in terms of the *ca*-rmsd, among each of the NMR-determined models and the corresponding X-ray structure for all these proteins was performed. From these analyses we can conclude the following: There are three proteins (see Table 1), Interleukin 13, the hyaluronan-binding domain of CD44 and the N-terminal domain of human tissue inhibitor of metalloproteinases-1, for which the *ca*-rmsd from the NMR-derived ensembles (1IK0, 2I83 and 1D2B) is similar or slightly better than the X-ray determined structures (3BPO, 1UUH and 2J0T); there are two other proteins, MIA protein and antifreeze protein CfAFP-501, for which the opposite is true, i.e., the rmsd of the X-ray structure (1I1J and 1M8N) is better than the *ca*-rmsd derived from the NMR ensemble of conformations (1HJD and 1Z2F). However, there is always at least one NMR-determined protein model for which the agreement between computed and observed $^{13}C^\alpha$ chemical shifts for cysteines in cystines is better than for the X-ray structure model (see for example Fig. S1 in Supplementary Material).

The analysis presented above must be complemented with a detailed analysis of the errors between observed and computed $^{13}C^\alpha$ chemical shifts, their origin and, if possible, a search for local flaws in the sequence, i.e., those that might reveal the need for further global or local refinement. Such analysis will be discussed in the next sub-sections.

Analysis of the errors

The frequencies of the error per-residue, Δ_μ^α, for all the cysteines in cystines, i.e., for 801 cysteine residues of cystines of all proteins, except BPTI, listed in Table 1, can be fit to a Gaussian distribution, with a mean value $x_0 = 0.12$ ppm, and standard deviation $\sigma = 3.69$ ppm (see Fig. 2). The resulting mean value ($x_o = 0.12$ ppm) is very close to the ideal one ($x_o = 0.0$ ppm) indicating that there is no need for further reference corrections, although the standard deviation ($\sigma = 3.69$ ppm) is significantly higher than the standard deviation ($\sigma = 1.64$ ppm) observed by Wang and Jardetzky (2002) for all cysteine residues of cystines. The observed standard deviation of $\sigma = 1.64$ ppm for the $^{13}C^\alpha$ chemical shifts pertains to cystines in *only* the β-strand conformation [the values from residues in statistical-coil and α-helix conformations were not included because the Wang and Jardetzky (2002) database does not contain enough statistics for oxidized cysteine residues]. Despite this, the computed high standard deviation ($\sigma = 3.69$ ppm) from 801 cystine residues of cysteine signals the following two possible problems: either the method is not accurate enough or most of the six protein structures used for the test (that does not include the BPTI models) need global or local refinement.

In order to determine whether the method is the origin of this problem, the following test was carried out. Six X-ray-determined BPTI structures, solved at 1.7 Å, or better, resolution (see Table 1), with low B-factors, namely for PDB id 1BPI, 1D0D, 1G6X, 1K6U, 5PTI and 6PTI, were used to compute the $^{13}C^\alpha$ chemical shift of each of the six cystines in each structure. The results indicate that the per-residue errors, except for all cysteines at position 14 and one cysteine (from 5PTI) at position 55 have values of $\Delta_\mu^\alpha \leq 1.64$ (data not shown). A brief analysis of the cysteines at position 14 for *all* the BPTI models will be presented below in a separate sub-section. In general, the average error over all cysteine residues, after excluding the six cysteines at position 14, is only 0.73 ppm, with a standard deviation of $\sigma = 0.47$ ppm. This standard deviation is ~8 times lower than the one obtained from the analysis of the 801 cystine residues ($\sigma = 3.69$ ppm) and within the observed standard deviation ($\sigma = 1.64$ ppm) obtained by Wang and Jardetzky (2002). This result enables us to rule out the method as the main source of the errors. These results also indicate that further refinement of the set of seven proteins, solved by both NMR spectroscopy and X-ray crystallography, might be necessary. In order to investigate this assumption, in the following section, we validate, first, the NMR- and, second, the X-ray-determined proteins, in terms of the *ca*-rmsd as the scoring function. Special attention to the factors that could contribute to the computed high standard deviation ($\sigma = 3.69$ ppm) will be explored.

Validation of the NMR-derived proteins

If the global quality of all the NMR-derived proteins listed in Table 1 were the main source of the high computed-standard deviation, then these proteins, as a whole, i.e., considering all non-cysteine residues, should have a higher *ca*-rmsd$_{76}$ value, computed by using Eq. 1, than a given cutoff, namely 2.6 ppm (see "Materials and methods" section). In column six of Table 1, we listed both the *ca*-rmsd and the *ca*-rmsd$_{76}$ value (in parentheses) for all non-cysteine residues. A comparative analysis against the chosen cutoff value indicates that three of the NMR-determined proteins show *ca*-rmsd$_{76}$ values greater than 2.6 ppm, namely proteins 1Z2F, 1HJD, and 1HA8, respectively, and the remaining three, namely proteins 2I83, 1IKO, and 1D2B, respectively, show a lower *ca*-rmsd$_{76}$ value. This result, by itself, does not enable us to reach any conclusive evidence indicating whether the global quality of the NMR-determined structures is the main origin of the high computed value for the standard deviations.

Does the above result imply that a local refinement, i.e., for only the cysteine residues of cystine, of the NMR-determined structures might be necessary? In order to answer this question, the *ca*-rmsd per-cysteine residue (shown in parentheses in column five, Table 1) for non-BPTI models was compared with the average *ca*-rmsd per-cysteine residue computed from the six BPTI models, after excluding Cys14 of the BPTI models (for the reasons explained below). The average *ca*-rmsd per-cysteine from the six BPTI models (0.95 ppm) is between ~3 and ~6 times lower than the *ca*-rmsd per-cysteine from any NMR-determine structure listed in Table 1 and, hence, indicates that the cysteine residues of cystine, in fact, must be locally refined.

Validation of the X-ray-derived proteins

In order to provide some insight into the most significant differences between observed and computed $^{13}C^{\alpha}$ chemical shifts computed for the X-ray determined structures, in the next sub-section we present a detailed analysis for the Cys14 of BPTI.

As possible sources of errors, the influence of the B-factors and the experimental conditions under which the X-ray and NMR experiments were carried out are discussed for two proteins, namely for Interleukin 13 and Melanoma inhibitory activity protein, in the Supplementary Material.

Residue Cys14 of the X-ray-solved models of BPTI

A detailed analysis of *all* the cysteines of cystine at position 14 in the sequence of the six BPTI X-ray-derived models (listed in Table 1) reveals an average error, $\Delta_{\mu}^{\alpha}=4.73$, and a standard deviation of $\sigma = 1.56$ ppm (data not shown). These values should be compared with the average value and standard deviation obtained for all cysteine residues of cystine after excluding the six cysteines at position 14, namely 0.73 and 0.47 ppm, respectively (data not shown).

Because disulfide bonds are observed to have several degrees of freedom (Van Wart and Scheraga 1976, 1977), disulfide-bond dynamics undergo significant conformational changes in solution (Otting et al. 1993; Sharma and Rajarathnam 2000). In this connection Otting et al. (1993) carried out a detailed NMR analysis of disulfide-bond isomerization in BPTI and in BPTI (G36S), a mutant protein with Gly replaced by Ser at position 36. Among other important findings, the authors found a slow dynamic equilibrium between two conformers with different chirality of the disulfide bond formed by Cys14 and Cys38, indicating that internal mobility prevails in this part of the molecule. Overall, flipped disulfide bonds may occur frequently in proteins in solution (as in BPTI) despite the conformational restraints imposed by the three dimensional structure (Otting et al. 1993) and, conceivably, this could be the origin of the significant difference between the observed (in solution) and computed (in a crystal) $^{13}C^{\alpha}$ chemical shifts of Cys14.

Analysis of the $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ chemical shifts as function of the redox state

As is well known (Sharma and Rajarathnam 2000), the redox state of cysteine residues can be straightforwardly inferred from $^{13}C^{\beta}$, but not from $^{13}C^{\alpha}$, chemical shifts. This conclusion was obtained by Sharma and Rajarathnam (2000) after statistical analysis of data from 375 $^{13}C^{\alpha}$ and 337 $^{13}C^{\beta}$ resonances from 132 proteins.

In the same computations of the $^{13}C^{\alpha}$ shielding of a given residue by the DFT methodology, we also obtain the shielding value of *all* nuclei in the residue, not only for the $^{13}C^{\alpha}$ nucleus; among them, also for the $^{13}C^{\beta}$ atom. Thus, we can investigate whether our theoretical calculations can reproduce the observed redox-induced behavior (Sharma and Rajarathnam 2000) of both the $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ chemical shifts. Hence, the $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ chemical shifts for the cysteines in cystine in both oxidation states for all proteins listed in Table 1 were obtained for each cysteine in both tripeptides and hexapeptides in the conformation of the regularized

experimental protein structure. The distributions of the computed $^{13}C^\alpha$ and $^{13}C^\beta$ chemical shifts for the cysteine residues in cystine and for reduced cysteines for all proteins listed in Table 1 are shown in Fig. 3 and Fig. 4, respectively. As shown in Fig. 3 and Fig. 4, these distributions can be modeled by a Gaussian function with a mean, x_0, value and standard deviation, σ. These Figures also show the existence of two, almost non-overlapped, basins for the $^{13}C^\beta$ (see Fig. 4), but not for the $^{13}C^\alpha$ (see Fig. 3), chemical shifts. These results are in good agreement with the observed $^{13}C^\alpha$ and $^{13}C^\beta$ chemical-shift values for cysteine residues in cystine and cysteine residues (Sharma and Rajarathnam 2000). As to whether such good agreement is quantitative or only qualitative follows.

Using the values that characterize the Gaussian distribution, shown in Fig. 3 and Fig.4, enables us to make a straightforward comparison with the observed (Sharma and Rajarathnam 2000) redox-induced shift effects. The values obtained for the computed $^{13}C^\alpha$ chemical shifts of cysteines in cystines, $x_0 = 56.6$ ppm, and $\sigma = 3.6$ ppm (see Fig. 3A), and reduced cysteines, $x_0 = 58.1$ ppm, and $\sigma = 4.2$ ppm (see Fig. 3B), are in good agreement with the observed (Sharma and Rajarathnam 2000) $^{13}C^\alpha$ chemical shifts values, $x_o = 55.5$ ppm, and $\sigma = 2.5$ ppm, and $x_o = 59.3$ ppm and $\sigma = 3.2$ ppm, respectively. A similar conclusion pertains to the computed $^{13}C^\beta$ chemical shifts of cysteines in cystines, $x_0 = 42.1$ ppm, and $\sigma = 5.8$ ppm (see Fig. 4A), and reduced cysteines, $x_0 = 27.3$ ppm, and $\sigma = 2.4$ ppm (see Fig. 4B), which also show good agreement with the observed values, $x_o = 40.7$ ppm, and $\sigma = 3.8$ ppm, and $x_o = 28.3$ ppm and $\sigma = 2.2$ ppm, respectively (Sharma and Rajarathnam 2000). Overall, our theoretical calculations are in good quantitative agreement with the observed $^{13}C^\alpha$ and $^{13}C^\beta$ chemical-shift values for reduced and oxidized cysteine residues (Sharma and Rajarathnam 2000).

These results on oxidized and reduced cysteines raise the question as to whether the computed $^{13}C^\alpha$ chemical shifts from reduced cysteines can be inferred from the values obtained from the cysteines in cystine, and vice versa. A visual inspection of Fig. 3 seems to indicate that the Gaussian distribution of the cysteines in cystine (Fig. 3A) is shifted by ~3 ppm with respect to the Gaussian distribution of the reduced cysteines (Fig. 3B). Does this observation imply that the computed $^{13}C^\alpha$ shielding for a given reduced cysteine can be obtained by applying a constant shift to the computed value from the oxidized state? The answer is no, for the following reason. The computed downfield shielding for reduced cysteine, with respect to the cysteines in cystine, is not equal for all cysteine residues nor do *all* cysteines show such a downfield shielding. In fact, ~33% of all 837 cysteines in cystine show upfield, rather than downfield, shielding (data not shown). This result indicates that explicit consideration of the disulfide bonds is a necessary condition for an accurate prediction of $^{13}C^\alpha$ chemical shifts of cysteines in cystines. In other words, the computed $^{13}C^\alpha$ chemical shifts from oxidized cysteines cannot be inferred straightforwardly from the values computed for the reduced state.

The above results are linked to the predictions of the recently introduced (Vila et al. 2009) $^{13}C^\alpha$ chemical shift (*Che*Shift) server because the predictions of the $^{13}C^\alpha$ chemical shift of *Che*Shift are valid *only* for reduced cysteine, and not for cysteine residues in cystine. Given that numerous proteins contain a large number of cysteine residues in cystine, a solution to this important problem is under investigation in our research group, and the results will be published elsewhere.

Conclusions

In this work, we present a method to compute, accurately, the $^{13}C^\alpha$ chemical shifts for cysteines in cystine. This new method has been applied to a selected, non-redundant, set of protein models, rich in disulfide bonds, determined by both NMR spectroscopy and X-ray crystallography. In particular, the analysis of a set of high-quality, X-ray-determined, protein models of BPTI enables us to both show the accuracy of the method to compute $^{13}C^\alpha$ chemical

shifts for the cysteines in cystine and to rule out the proposed DFT-computational methodology as the main source of the computed errors in the chosen set of proteins. Thus, the errors between computed and observed $^{13}C^{\alpha}$ chemical shifts for the cysteines of cystine originated in several factors that include, but are not limited to, the need for further refinement of NMR-determined conformations, the presence of high B-factors, as for X-ray-determined conformations, or poor representation of the disulfide-bond dynamics in solution by a single conformation, etc.

By using quantum-chemical-based calculations we have been able to illustrate that $^{13}C^{\beta}$, but not $^{13}C^{\alpha}$, chemical shifts (see Fig. 3, Fig. 4) show two, almost non-overlapped, basins, in good agreement with the observation of Sharma and Rajarathnam (2000) and, hence, providing a validation of the methodology used here to compute $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ chemical shifts of cysteines as a function of the redox state. Moreover, we have been able to demonstrate that the disulfide bond significantly affects the computed values of the $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ chemical shifts and, hence, explicit consideration of the presence of a disulfide bond is necessary for an accurate prediction of $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ chemical shifts of cysteines in cystines.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

Arnautova YA, Vila JA, Martin OA, Scheraga HA. Acta Cryst D 2009;D65:697–703. [PubMed: 19564690]

Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The protein data bank. Nucleic Acids Res 2000;28:235–242. [PubMed: 10592235]

Betancourt MR, Skolnick J. Universal similarity measure for comparing protein structures. Biopolymers 2001;59:305–309. [PubMed: 11514933]

Carugo O, Pongor S. A normalized root-mean-square distance for comparing protein three-dimensional structures. Proteins Sci 2001;10:1470–1473.

Chesnut DB, Moore KD. Locally dense basis-sets for chemical-shift calculations. J Comp Chem 1989;10:648–659.

Cornilescu G, Marquardt JL, Ottiger M, Bax A. Validation of protein structure from anisotropic carbonyl chemical shifts in a dilute liquid crystalline phase. J Am Chem Soc 1998;120:6836–6837.

de Dios AC, Pearson JG, Oldfield E. Secondary and tertiary structural effects on protein NMR chemical shifts: an ab initio approach. Science 1993;260:1491–1496. [PubMed: 8502992]

Frisch, MJ.; Trucks, GW.; Schlegel, HB.; Scuseria, GE.; Robb, MA.; Cheeseman, JR.; Zakrzewski, VG.; Montgomery, JA.; Stratmann, RE., Jr; Burant, JC.; Dapprich, S.; Millam, JM.; Daniels, AD.; Kudin, KN.; Strain, MC.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, GA.; Ayala, PY.; Cui, Q.; Morokuma, K.; Malick, DK.; Rabuck, AD.; Raghavachari, K.; Foresman, JB.; Cioslowski, J.; Ortiz, V.; Baboul, AG.; Stefanov, BB.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, RL.; Fox, DJ.; Keith, T.; Al-Laham, MA.; Peng, CY.; Nanayakkara, A.; Gonzalez, C.; Chal-lacombe, M.; Gill, PMW.; Johnson, B.; Chen, W.; Wong, MW.; Andres, JL.; Gonzalez, C.; Head-Gordon, M.; Replogle, ES.; Pople, JA. Gaussian 03, Revision E.01. Wallingford: Gaussian; 2004.

Havlin RH, Le H, Laws DD, de Dios AC, Oldfield E. An ab initio quantum chemical investigation of carbon-13 NMR shielding tensors in glycine, alanine, valine, isoleucine, serine, and threonine:

comparisons between helical and sheet tensors, and effects of χ1 on shielding. J Am Chem Soc 1997;119:11951–11958.

Iyer S, Wei S, Brew K, Acharya KR. Crystal structure of the catalytic domain of matrix metalloproteinase-1 in complex with the inhibitory domain of tissue inhibitor of metalloproteinase-1. J Biol Chem 2007;282:364–371. [PubMed: 17050530]

Kornhaber GJ, Snyder D, Moseley HNB, Montelione GT. Identification of zinc-ligated cysteine residues based on chemical shift data $^{13}C^\alpha$ and $^{13}C^\beta$. J Biomol NMR 2006;34:259–269. [PubMed: 16645816]

Li C, Guo X, Jia Z, Xia B, Jin C. Solution structure of an antifreeze protein CfAFP-501 from choristoneura fumiferana. J Biomol NMR 2005;32:251–256. [PubMed: 16132825]

Maiorov VN, Crippen GM. Size-independent comparison of protein three-dimensional structures. Proteins 1995;22:273–283. [PubMed: 7479700]

Otting G, Liepinsh E, Wüthrich K. Disulfide bond isomerization in BPTI and BPTI(G36S): an NMR study of correlated mobility in proteins. Biochemistry 1993;32:3571–3582. [PubMed: 7682109]

Pearson JG, Le H, Sanders LK, Godbout N, Havlin RH, Oldfield EJ. Predicting chemical shifts in proteins: structure refinement of valine residues by using ab initio and empirical geometry optimizations. J Am Chem Soc 1997;119:11941–11950.

Sharma D, Rajarathnam K. ^{13}C NMR chemical shifts can predict disulfide bond formation. J Biomol NMR 2000;18:165–171. [PubMed: 11101221]

Spera S, Bax A. Empirical correlation between protein backbone conformation and C^α and $C^{\beta}$$^{13}C$ nuclear magnetic resonance chemical shifts. J Am Chem Soc 1991;113:5490–5492.

Sun H, Sanders LK, Oldfield E. Carbon-13 NMR shielding in the twenty common amino acids: comparisons with experimental results in proteins. J Am Chem Soc 2002;124:5486–5495. [PubMed: 11996591]

Ulrich EL, Akutsu H, Doreleijers HJ, Harano Y, Ioannidis YE, Lin J, Livny M, Mading S, Maziuk D, Miller Z, Nakatani E, Schulte CF, Tolmie DE, Wenger RK, Yao H, Markley JL. BioMagResBank. Nucleic Acids Res 2007;36:D402–D408. [PubMed: 17984079]

Van Wart HE, Scheraga HA. Raman spectra of cystine-related disulfides. Effect of rotational isomerism about carbon-sulfur bonds on sulfur-sulfur stretching frequencies. J Phys Chem 1976;80:1812–1823.

Van Wart HE, Scheraga HA. Stable conformations of aliphatic disulfides: influence of 1, 4 interactions involving sulfur atoms. Proc Natl Acad Sci USA 1977;74:13–17. [PubMed: 264666]

Vila JA, Scheraga HA. Factors affecting the use of $^{13}C^\alpha$ chemical shifts to determine, refine, and validate protein structures. Proteins: Struct Funct Bioinform 2008;71:641–654.

Vila JA, Scheraga HA. Assessing the accuracy of protein structures by quantum mechanical computations of $^{13}C^\alpha$ chemical shifts. Acc Chem Res 2009;42:1545–1553. [PubMed: 19572703]

Vila JA, Villegas ME, Baldoni HA, Scheraga HA. Predicting $^{13}C^\alpha$ chemical shifts for validation of protein structures. J Biomol NMR 2007;38:221–235. [PubMed: 17558470]

Vila JA, Aramini JA, Rossi P, Kuzin A, Su M, Seetharaman J, Xiao R, Tong L, Montelione GT, Scheraga HA. Quantum chemical $^{13}C^\alpha$ chemical shift calculations for protein NMR structure determination, refinement, and validation. Proc Natl Acad Sci USA 2008;105:14389–14394. [PubMed: 18787110]

Vila JA, Arnautova YA, Martin OA, Scheraga HA. Quantummechanics-derived $^{13}C^\alpha$ chemical shift server (CheShift) for protein structure validation. Proc Natl Acad Sci USA 2009;106:16972–16977. [PubMed: 19805131]

Villegas ME, Vila JA, Scheraga HA. Effects of side-chain orientation on the $^{13}C^\alpha$ chemical shifts of antiparallel β-sheet model peptides. J Biomol NMR 2007;37:137–146. [PubMed: 17180547]

Wang Y, Jardetzky O. Probability-based protein secondary structure identification using combined NMR chemical-shift data. Protein Sci 2002;11:852–861. [PubMed: 11910028]

Wouters MA, George RA, Haworth NL. "Forbidden" disulfides: their role as redox switches. Curr Protein Pept Sci 2007;8(5):484–495. [PubMed: 17979763]

Xu XP, Case DAJ. Automatic prediction of ^{15}N, $^{13}C^\alpha$, $^{13}C^\beta$ and $^{13}C^\chi$ chemical shifts in proteins using a density functional database. J Biomol NMR 2001;21:321–333. [PubMed: 11824752]

Fig. 1.
Ribbon diagram of the hexapeptide model used to compute the $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ chemical shifts for cysteines in disulfide bonds. The hexapeptide is formed by two tripeptides Ac-G**C$_i$**G-NMe and Ac-G**C$_j$**G-NMe, with **C$_i$** *and* **C$_j$** and the disulfide bond in the conformation of the regularized experimental protein structure, and the Gly, Ac and NMe residues in their energy-minimized structure. The Cys residues with their disulfide bond are colored in orange, the Gly residues are colored white and the blocking end-groups are red and blue for Ac and NMe, respectively
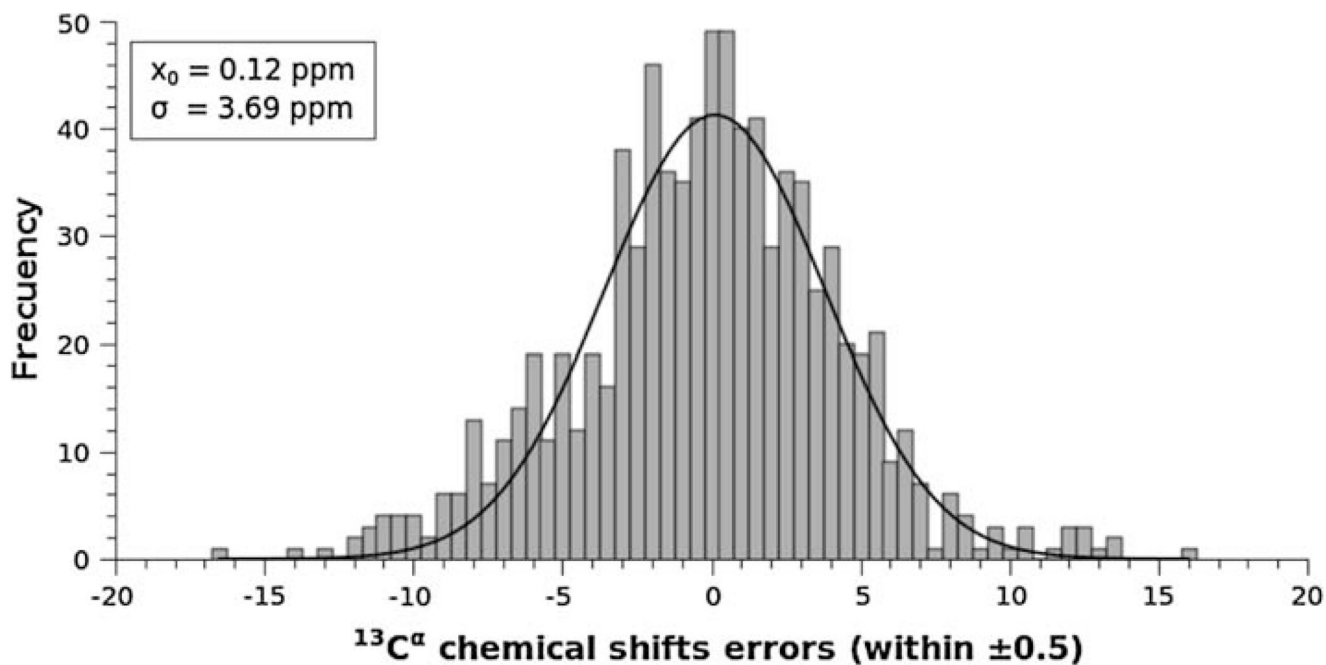
Fig. 2.
Frequency of the error, Δ_μ^α, between computed and observed $^{13}C^\alpha$ chemical shifts, for each of the 801 cysteine residues in cystines computed in this work, i.e., from all proteins listed in Table 1, except BPTI. The data were binned using an interval of 0.5 ppm, and then fit to a Gaussian function with a mean value $x_0 = 0.12$ ppm, and standard deviation $\sigma = 3.69$ ppm
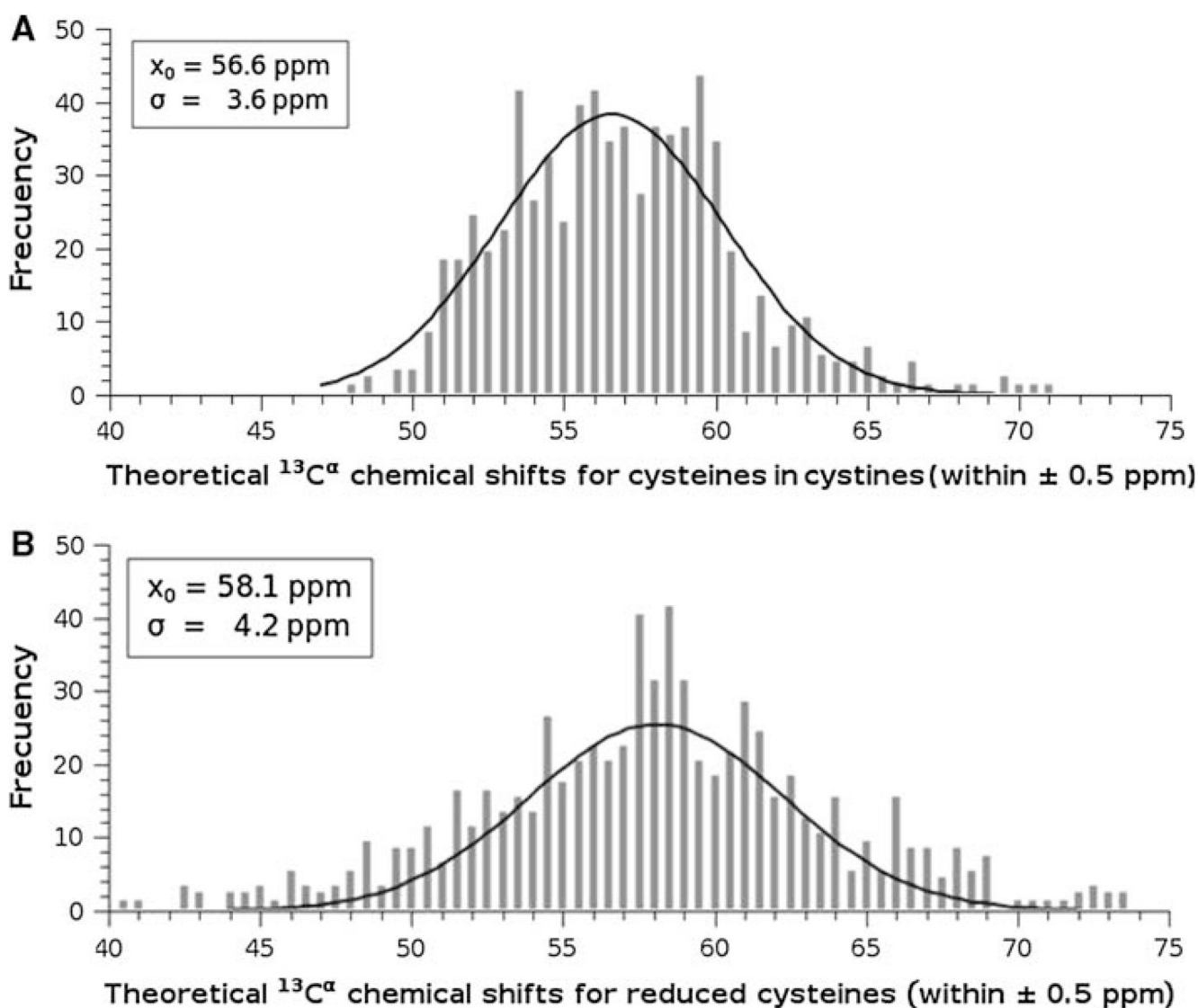
Fig. 3.
Frequency of the computed $^{13}C^\alpha$ chemical shifts for all 837 cysteines in cystine: **A** for cysteines in cystine. The data were binned using an interval of 0.5 ppm, and then fitted to a Gaussian function with a mean value $x_0 = 56.6$ ppm, and standard deviation of $\sigma = 3.6$ ppm; **B** same as **A** for reduced cysteines; with $x_0 = 58.1$ ppm, and standard deviation of $\sigma = 4.2$ ppm
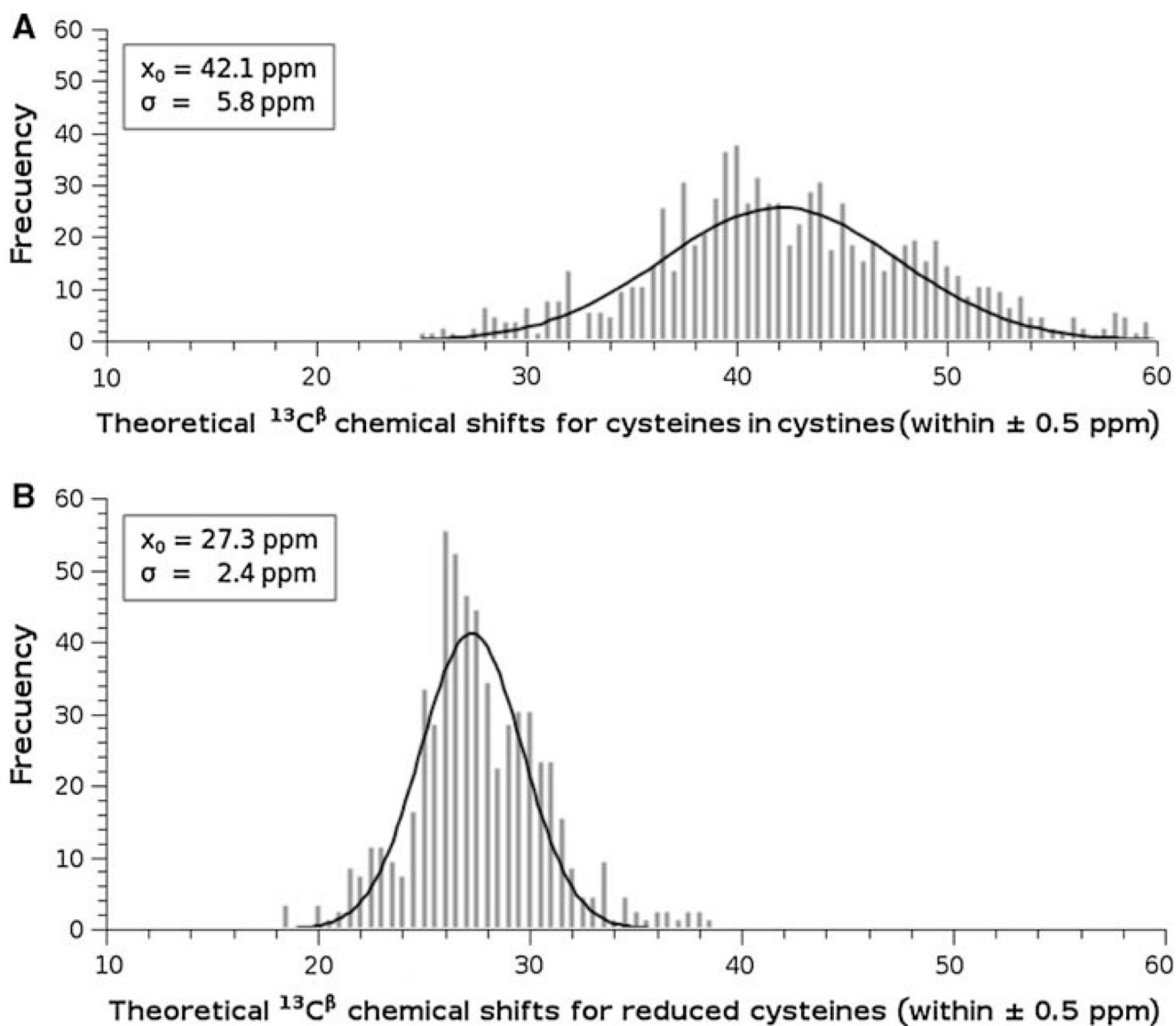
Fig. 4.
Frequency of the computed $^{13}C^\beta$ chemical shifts for all 837 cysteines in cystine: **A** for cysteines in cystine. The data were binned using an interval of 0.5 ppm, and then fitted to a Gaussian function with a mean value $x_0 = 42.1$ ppm, and standard deviation $\sigma = 5.8$ ppm; **B** same as **A** for reduced cysteines; with $x_0 = 27.3$ ppm, and standard deviation $\sigma = 2.4$ ppm

Table 1

Set of proteins for which the $^{13}C^{\alpha}$ chemical shift was computed

PDB code[a]	Experimental conditions[b]	Number of residues[c]	BMRB accession code[d]	ca-rmsd[e] (ppm)	ca-rmsd[f] (ppm)
1Z2F (20)[g]	NMR (N/A; 290 K; 5.7)	121 (9)	[6111]	3.5	4.25 (3.56)
1M8N (1)[g]	X-ray (2.45 Å; 100 K; 5.2)	120 (9)		3.2	4.38 (3.68)
2I83 (20)[h]	NMR (N/A; 310 K; 6.7)	158 (6)	[5903]	2.45	2.21 (1.64)
1UUH (1)[h]	X-ray (2.2 Å; 100 K; 5.5)	150 (6)		2.47	3.1 (2.35)
1HJD (20)	NMR (N/A; 300 K; 7.0)	101(2)	[4731]	6.08	4.13 (3.65)
1I1J (1)	X-ray (1.39 Å; 100 K; 8.2)	101 (2)		5.16	3.81 (3.39)
1IK0 (30)	NMR (N/A; 298 K; 6.0)	113 (4)	[5004]	4.92	2.37 (2.01)
3BPO (1)	X-ray (3.00 Å; 100 K; 6.0)	98 (4)		5.21	2.9 (2.62)
1D2B (29)[i]	NMR (N/A; 293 K; 6.0)	119 (4)	[4327]	2.61	2.74 (2.27)
2I0T (1)[i]	X-ray (2.54 Å; 100 K; 7.5)	119 (4)		2.56	3.43 (2.84)
1BPI (1)	X-ray (1.09 Å; 125 K; N/A)	58 (6)	[5359]	1.67 [0.86]	1.65 (2.03)
1D0D (1)	X-ray (1.62 Å; 298 K; 6.5)	58 (6)		1.33 [1.09]	1.94 (2.39)
1G6X (1)	X-ray (0.86 Å; 100 K; 7.5)	58 (6)		2.39 [1.05]	1.78 (2.20)
1K6U (1)	X-ray (1.00 Å; 100 K; 7.5)	58 (6)		2.39 [0.97]	2.02 (2.49)
5PTI (1)	X-ray (1.09 Å; N/A; N/A)	57 (6)		1.66 [1.08]	1.69 (2.11)
6PTI (1)[j]	X-ray (1.7 Å; N/A; N/A)	56 (6)		2.37 [0.64]	1.81 (2.29)
1HA8 (20)	NMR (N/A; 290 K; 4.6)	51 (10)	[4979]	4.42	3.16 (4.57)

[a] Four-symbol code for the deposited structure in the Protein Data Bank (Berman et al. 2000). In parentheses, the number of determined conformations for each protein

[b] Experimental conditions under which the proteins listed in column one were determined. The resolution (Å), temperature (K) and pH, are in parentheses. For all the NMR structures, DSS was used as the reference for the observed $^{13}C^{\alpha}$ chemical shifts

[c] Total number of residues for each protein listed in column one. The number of cysteine residues in cystine for each protein is indicated between parentheses. For the same protein, and for three out of five cases, the total number of residues of each structure solved by NMR spectroscopy and X-ray crystallography are not the same; the reasons for the observed experimental difference in the total number of residues, can be found in the original papers cited in "Materials and methods" section

[d] BMRB (Ulrich et al. 2007) accession number under which the observed $^{13}C^\alpha$ chemical shifts can be found

[e] ca-rmsd computed from only the cysteine residues in cystines. The ca-rmsd values without cysteine in position 14, only for the six BPTI models, are shown in brackets

[f] ca-rmsd computed for non-cysteine residues; in parenthesis, the normalized size-independent ca-rmsd76

[g] The all heavy-atom rmsd value between the (1Z2F) NMR- and the X-ray-determined (1M8N) structure is ~0.65 Å. Differences between the 1Z2F and 1M8N models are located mainly in two loop regions (Li et al. 2005), namely for residues 90–93 and 106–110, near the C-terminal region. The X-ray-determined structure (1M8N) was solved as a dimer, but such oligomerization is not observed in solution and, hence, the NMR-determined structure (1Z2F) was solved as a monomer (Li et al. 2005). There is an odd number (nine) of cysteines listed in parentheses in column three because the observed $^{13}C^\alpha$ chemical-shift value for one of the cysteines is missing

[h] The X-ray determined model (1UUH) was solved without ligand bound, while the NMR-determined structure (2I83) was solved with bound ligand. Several regions of the Hyaluronan-binding domain of the CD 44 protein undergo conformational changes upon ligand binding, as reflected by a high (~7.7 Å) all-heavy-atom rmsd between the 1UUH and 2I83 protein models. Most of the conformational changes occur at the C-terminal portion, i.e., for the last ~39 residues of the protein. Excluding the C-terminal portion from the rmsd analysis, a value of ~1.8 Å is obtained, indicating that there are still significant conformational differences between the 1UUH and 2I83 protein models

[i] The rmsd between the X-ray model (2J0T) and the average NMR-determined conformation (1D2B) is 1.49 Å (Iyer et al. 2007). Conceivably, most of the conformational differences between these models arise from the fact that the NMR-derived structure (1D2B) was solved as a monomer while the X-ray-determined protein (2J0T) was solved as a dimer

[j] Protein refined by neutron diffraction data