Supplementary Material

Analysis of $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ Chemical Shifts of Cysteine and Cystine Residues in Proteins: A Quantum Chemical Approach

Osvaldo A. Martin,[*] Myriam E. Villegas,[*] Jorge A.Vila,[†*] and Harold A. Scheraga[†‡]

[*] Universidad Nacional de San Luis, Instituto de Matemática Aplicada 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.

[‡]corresponding author: has5@cornell.edu

Experimental set of structures. The structures of five of the seven structures listed in Table 1 of the main text were determined experimentally by both NMR spectroscopy and X-ray crystallography, namely (*1*) *Choristoneura fumiferana* Antifreeze Protein, Isoform 501; PDB id 1Z2F (Li *et al,* 2005) and PDB id 1M8N chain A (Leinala *et al,* 2002); (*2*) Hyaluronan-binding domain of CD 44; PDB id 2I83 (Takeda *et al,* 2006) and PDB id 1UUH chain A (Teriete *et al,* 2004); (*3*) Melanoma inhibitory activity (MIA) protein; PDB id 1HJD (Stoll *et al,* 2001) and PDB id 1I1J chain A (Lougheed *et al,* 2001); (*4*) Human Inter Leukine 13; PDB id 1IK0 (Moy *et al,* 2001) and PDB- id 3BPO chain A (Laporte *et al,* 2008); and (*5*) the MMP-inhibitory, N-terminal domain of human tissue inhibitor of metalloproteinases-1; PDB id 1D2B (Wu *et al*, 2000) and PDB id 2J0T Chain A (Iyer *et al*, 2007). The structure of the sixth protein, namely Bovine Pancreatic Trypsin Inhibitor**;** PDB-id: 1BPI, 1D0D chain A, 1G6X, 1K6U, 5PTI and 6PTI (Parkin *et al,* 1996; Charles *et al*, 2000; Addlagatta *et al,* 2001; Botos *et al,* 2001; and Wlodawer *et al*, 1984, 1987, respectively), was determined only by X-ray crystallography. The structure of the seventh protein, Pheromone ER-23 PDB id 1HA8, was determined only by NMR-spectroscopy (Zahn et al 2001).

Six of these proteins were selected by cross-referencing the Protein Data Bank (PDB) [Berman *et al*, 2000] with the Biological Magnetic Resonance Data Bank (BMRB) [Ulrich *et al*, 2007] in order to obtain a non-redundant set of structures, i.e., a set of proteins possessing sequence identity below 30%. Additionally, 6 high-resolution X-ray protein models of BPTI were chosen for further analysis. With this set of 7 proteins, the total number of cysteines, for which the $^{13}C^{\alpha}$ chemical shifts were computed at the DFT level of theory is 837, from NMR- and X-ray-

determined conformations, respectively; the same number of $^{13}C^{\beta}$ chemical shifts for cysteine residues in both the oxidized and reduced state was also computed at the DFT level of theory.

There is an odd number (837) of cysteine residues because one of the observed $^{13}C^{\alpha}$ chemical-shift values of the oxidized cysteines is missing from the *Choristoneura fumiferana* Antifreeze Protein Isoform 501 (PDB code 1Z2F and 1M8N).

Method to compute $^{13}C^{\alpha}$ chemical shifts. All the experimentally determined conformations were first *regularized,* i.e., all residues were replaced by the standard ECEPP/3 (Némethy et al, 1992) residue geometry in which bond lengths and bond angles are fixed (rigid-body geometry approximation) at standard values, and hydrogen atoms are added, if necessary. The final conformations resulting from the regularization procedure are close to the experimental structures for all cases, with an average rmsd value for all heavy atoms of ~ 0.17 Å.

The $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ chemical shifts for each amino acid residue $\mathbf{C}_i$ and $\mathbf{C}_j$ were computed at the OB98/6-311+G(2d,p) level of theory, while the remaining residues in the hexapeptide were treated at the OB98/3-21G level of theory, i.e., by using the *locally-dense* approach (Chesnut & Moore, 1989); (*c*) the computed $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ shieldings ($\sigma^{n}_{subst,th}$, with $n = \alpha$ or β) were converted to $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ chemical shifts (δ^{n}) by employing the equation $\delta^{n}_{th} = \sigma^{n}_{ref} - \sigma^{n}_{subst,th}$, where the indices denote a theoretical (*th*) computation, the reference substance (*ref*), and the substance of interest (*subst*), i.e., the $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ shielding, respectively, of a given amino acid residue $\mathbf{C}_i$ and $\mathbf{C}_j$.

All the computed $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ shielding values ($\sigma^{n}_{subst,th}$, with $n = \alpha$ or β) 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). We have used only one

exchange-correlation functional, namely OB98, because it was shown that this functional is, among others, one of the most accurate *and* faster ones with which to reproduce the observed $^{13}C^{\alpha}$ chemical shifts of proteins in solution (Vila et al., 2009).

All the calculated isotropic shielding values were referenced with respect to a tetramethylsilane (TMS) chemical shift scale, as described previously (Vila et al. 2002; Vila et al. 2009). All the experimental chemical shift data used in this work were obtained by using 2, 2-dimethyl-2-silapentane-5-sulfonic acid (DSS) as the reference compound. Conversion of the computed TMS-referenced values for the $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ shielding chemical shifts to DSS was carried out by adding 1.70 ppm to the computed values (Wishart et al., 1995).

Determination of an effective TMS shielding value for $^{13}C^{\beta}$. By adopting the observed TMS value of 188.1 ppm (Jameson & Jameson 1987), it is possible to find the characteristic mean (x_o) and standard deviation (σ) of the Gaussian function that fits the frequency of the error distribution per-residue (Δ_{μ}^{β}), for all 837 cysteine residues of cystines. The characteristic mean value (x_o) appears displaced from the ideal value of 0.0 by 8.96 ppm and, hence, use of an *effective* TMS value (Vila et al. 2009) of 179.1 ppm for C^{β} gives $x_o = 0.0$.

In a similar way (Vila et al. 2009), the *effective* TMS value used to compute the $^{13}C^{\alpha}$ chemical shifts (with $x_o = 0.0$) was 184.5 ppm. Since $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ are involved in different bonding arrangements, it is not surprising that different effective TMS values are obtained for each nucleus.

Computation of the conformationally-averaged rmsd (ca-rmsd). A protein in solution exists as

an ensemble of conformations and, hence, we can assume that the observed chemical shifts $^{13}C^{n}_{observed,\mu}$ with $n = \alpha$ or β, for a given amino acid residue μ can be interpreted as a conformational average over different rotational states represented by a discrete number of different conformations; all of these conformations are assumed to satisfy the NMR constraints from which the conformations were derived (Vila et al., 2007). Thus, the following quantity can be computed: $^{13}C^{n}_{computed,\mu} = \sum_{i=1}^{\Omega} \lambda_i \, ^{13}C^{n}_{\mu,i,}$ where $^{13}C^{n}_{\mu,i}$ is the computed chemical shift for amino acid μ in conformation i out of Ω protein conformations, and λ_i is the Boltzmann weight factor for conformation i, with the condition $\sum_{i=1}^{\Omega} \lambda_i \equiv 1$. With existing computational resources, it is not feasible to determine λ_i at the quantum chemical level, and, hence, it is assumed that the following equality is always valid: $\lambda_i \equiv 1/\Omega$. In other words, under conditions of fast conformational averaging, we assume that all Boltzmann weight factors contribute equally. Under these assumptions, the computation of the *ca*-rmsd for a protein containing N amino acids residues, is straightforward (Vila et al., 2007): $ca\text{-rmsd}^{n} = [(1/N) \sum_{\mu=1}^{N} (^{13}C^{n}_{observed,\mu} - <^{13}C^{n}_{computed}>_{\mu})^2]^{1/2}$ with $n = \alpha$ or β, and $<^{13}C^{n}_{computed}>_{\mu} = (1/\Omega) \sum_{i=1}^{\Omega} \, ^{13}C^{n}_{\mu,i}$. In addition, for each amino acid μ, we define an error function $\Delta^{n}_{\mu} = (^{13}C^{n}_{observed,\mu} - ^{13}C^{n}_{computed,\mu})$, with $n = \alpha$ or β.

Interleukin 13 protein (3BPO). The origin of the high computed *ca*-rmsd value (5.21 ppm, listed

in Table 1) for the cysteine residues of cystine of the X-ray-determined protein model of 3BPO may be due to both the high average B factors for the 2 cystines in the structure, i.e., with an average of 51.6 ± 5.48 $Å^2$, and the low resolution (3.0 Å) at which the structure was determined, e.g., in this structure, 29 out of 127 residues were missing in the electron density map, and 9 residues were missing heavy atoms.

Another possible source of error is the difference in the experimental conditions under which the X-ray and NMR experiments were carried out, namely, the X-ray structure was determined as a co-crystal of a ternary complex (Interleukin 13-Interleukin 4- Chain "A" of the Interleukin 13 Receptor) whereas the observed $^{13}C^{\alpha}$ chemical shifts were obtained from Interleukin 13 in the unbound state. The rmsd between the X-ray and NMR models of Interleukin 13 is ~1.18 Å on average.

Melanoma inhibitory activity (MIA) protein (1I1J). In contrast to the previous analysis for the Interleukin 13 protein, the X-ray-determined model for the MIA protein (1I1J) has been solved at high resolution (1.39 Å), and the average B-factors are significantly lower than those for the Interleukin 13 protein, namely 21.9 $Å^2$ for the whole protein and 23.32 $Å^2$ for all the cysteine residues. Additionally, only 2 residues are missing in the electron density map, i.e., the first and the last one. Despite this, the computed *ca*-rmsd value (5.16 ppm) for the cysteine residues of cystine is very similar to the one computed for the Interleukin 13 protein (5.21 ppm) as shown in Table 1, indicating that the high quality of the protein structure should not be the main origin of the computed errors.

Regarding possible source of errors, it is worth noting that ~49% of the residues of the MIA protein are located in non-regular portions of the molecule, indicating that the protein might

be very flexible in solution. If this were the case, the $^{13}C^{\alpha}$ chemical shift should reflect the dynamics of such structural elements. In other words, the low temperature at which the X-ray structure was determined, namely 100 K (although the NMR resonances were collected at 300 K), might have led to a crystal structure in which residues populating a flexible portion of the molecule, such as Cys17 that shows a $|\Delta_{\mu}^{\alpha}|$ = ~ 8.0 ppm, might not be good representations of the dynamics in solution.

A normalized rmsd for comparing different protein structures. Using different metrics, such as the rmsd, the $rmsd_L$, as given by Eq. (1) in the main text, and the rmsd-per residue, we carry out a comparison of the 'quality' of proteins of different sizes, in terms of the agreement between observed and predicted $^{13}C^{\alpha}$ chemical shifts of their structures. For this analysis, 24 NMR-derived proteins with PDB id: 1B22, 1B2T, 1B4R, 1BBN, 1BJX, 1BLR, 1BNO, 1BQZ, 1CK2, 1CZ4, 1DC2, 1DOQ, 1E0G, 1E17, 1EIG, 1EWW, 1EZA, 1EZO, 1F2H, 1F3Y, 1F43, 1FAF, 1D3Z, and 2JVD, not listed in Table 1 of the main text, were chosen. This selection covers a wide range of protein sizes from 48 to 370 residues. For each of these proteins, the rmsd between the observed values of the $^{13}C^{\alpha}$ chemical shifts, obtained from the BRMB database, and the predicted values obtained from the *Che*Shift server were computed.

Figure S2a shows the distribution of the rmsd's mentioned above. In particular, the value obtained for protein 2JVD (green-filled circle) is the rmsd which is among the four lowest values. This raises the question as to whether a protein with 48 residues (2JVD), and rmsd = 1.99 ppm, is of better quality than a protein, e.g., with 370 residues (1EZO), and rmsd = 2.72 ppm. This has been a long-lasting problem in the field of protein structure prediction and determination (Maiorov & Crippen, 1995; Betancourt & Skolnick, 2001; Carugo & Pongor, 2001). In other

words, the rmsd is a reliable indicator of the global property of protein structures only when containing the same, or similar, numbers of residues. The attempt to solve this problem has been provided by Carugo & Pongor (2001), among others. Thus, an analysis based on Equation (1) of the main text for the $rmsd_{76}$ (shown by grey-filled squares in Figure S2b) and the rmsd-per-residue (shown by red-filled triangles in Figure S2b) indicates that protein 2JVD possesses a normalized, size-independent, $rmsd_{76}$ higher than any protein with $N > 100$ residues. Additionally, for proteins containing less than 125 residues (see Figure S2b), the agreement in terms of the correlation coefficient R, between the rmsd-per-residue and the $rmsd_{76}$ is fairly good, namely $R = 0.98$, although the rmsd-per-residue *significantly* overestimates the quality for structures containing higher than ~100 residues (rmsd of 0.56 ppm for 370 residues protein compared to 1.77 ppm for 76 residues of ubiquitin, in Figure S2b) and, hence, the rmsd-per-residue is not a reliable metric.

The frequency of the distribution of the $rmsd_{76}$ values (the grey-filled squares in Figure S2b) is shown in Figure S3. These data can be fit by a Gaussian or Normal distribution with a mean value x_o ~2.0 ppm and a standard deviation $\sigma \sim 0.3$ ppm, as indicated in the inserted panel in Figure S3. Thus, an $rmsd_{76} = 2.6$ ppm, i.e., within $\sim 2\sigma$ of the mean $rmsd_{76}$ value, can be adopted as a cutoff value. In this way, we assume that ~80% of the structures, i.e., those possessing an $rmsd_{76} \leq 2.6$ ppm, have similar quality as the reference structure, namely model 1 of ubiquitin (1D3Z); the remaining ones, i.e., 5 out of 24 shown in Figure S2b, are assumed to need further refinement.

A further advantage of the use of Eq.(1) to compare quality of structures with different sizes is the following. The dispersion of the rmsd values shown in Figure S2a among the 24 proteins is ~1.75 ppm while the corresponding dispersion of the $rmsd_{76}$ values, shown in Figure

S2b, is more than twice and, hence, offers a more sensitive rmsd range of distribution with which to discriminate structures.

Finally, it is worth noting that the whole analysis in this section has been carried out by using a single structure, namely the first one, if more than one exists, as a representative structure of the ensemble of NMR-derived conformations. This procedure was adopted for simplicity, although, in practical applications, the *ca*-rmsd$_{76}$, rather than the rmsd$_{76}$, *must be* considered if more than one structure is available.

References

Addlagatta A, Krzywda S., Czapinska H, Otlewski J, Jaskolski M. (2001) Ultrahigh-resolution structure of a BPTI mutant. Acta Cryst, Sect.D 57: 649-663.

Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H,Shindyalov IN, Bourne PE (2000) The Protein Data Bank. Nucleic Acids Res 28:235-242.

Betancourt MR, Skolnick J (2001) Universal similarity measure for comparing protein structures. Biopolymers 59:305-309.

Botos I, Wu Z, Lu W, Wlodawer A (2001) Crystal structure of a cyclic form of bovine pancreatic trypsin inhibitor. FEBS Lett. 509: 90-94.

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

Charles R, Padmanabhan K, Arni RV, Padmanabhan KP, Tulinsky A. (2000) Structure of tick anticoagulant peptide at 1.6 A resolution complexed with bovine pancreatic trypsin inhibitor. Protein Sci. 9: 265-272.

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

Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA,Cheeseman JR, Zakrzewski VG, Montgomery JA, Stratmann REJr, 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, OrtizV, 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, AndresJL, Gonzalez C, Head-Gordon M, Replogle ES, Pople JA, Gaussian 03, Revision E.01, Gaussian, Inc., Wallingford CT, 2004

Iyer S, Wei S, Brew K, Acharya KR. (2007) 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, 282: 364-371

Jameson AK, Jameson CJ (1987) Gas-phase 13C chemical shifts in the zero-pressure limit: Refinements to the absolute shielding scale for 13C. J Chem Phys Lett, 134:461-466

Laporte SL, Juo ZS, Vaclavikova J, Colf LA, Qi X, Heller NM, Keegan AD, Garcia KC. (2008) Molecular and Structural Basis of Cytokine Receptor Pleiotropy in the Interleukin-4/13 System. Cell(Cambridge,Mass.)132: 259-272

Leinala EK, Davies PL, Doucet D, Tyshenko MG, Walker VK, Jia Z. (2002) A beta-helical antifreeze protein isoform with increased activity: structural and functional insights J.Biol.Chem. 277: 33349-33352

Li C, Guo X, Jia Z, Xia B, Jin C. (2005) Solution Structure of an Antifreeze Protein CfAFP-501 from Choristoneura fumiferana J.Biomol.Nmr 32: 251-256.

Lougheed JC, Holton JM, Alber T, Bazan JF, Handel TM (2001) Structure of melanoma inhibitory activity protein, a member of a recently identified family of secreted proteins. Proc Natl Acad Sci USA, 98: 5515-5520.

Maiorov VN, Crippen GM (1995) Size-independent comparison of protein three-dimensional structures. Proteins 22:273-283.

Moy FJ, Diblasio E, Wilhelm J, Powers R (2001) Solution structure of human IL-13 and implication for receptor binding. J Mol Biol, 310: 219-230

Némethy G, Gibson KD, Palmer KA, Yoon CN, Paterlini G, Zagari A, Rumsey S,

Scheraga HA (1992). Energy parameters in polypeptides. 10. Improved geometrical parameters and non-bonded interactions for use in the ECEPP/3 algorithm, with application to proline containing peptides. J Phys Chem 96:6472–6484.

Parkin S, Rupp B, Hope H (1996) Structure of bovine pancreatic trypsin inhibitor at 125 K definition of carboxyl-terminal residues Gly57 and Ala58. Acta Cryst, Sect D 52: 18-29

Stoll R, Renner C, Zweckstetter M, Bruggert M, Ambrosius D. Palme S. Engh RA, Golob M, Breibach I, Buettner R, Voelter W, Holak TA, Bosserhoff AK (2001) The Extracellular Human Melanoma Inhibitory Activity (MIA) Protein Adopts an SH3 Domain-Like Fold. Embo J, 20:340-349.

Takeda M, Ogino S, Umemoto R, Sakakura M, Kajiwara M, Sugahara KN, Hayasaka H, Miyasaka M, Terasawa H, Shimada I. (2006) Ligand-induced Structural Changes of the CD44 Hyaluronan-binding Domain Revealed by NMR J.Biol.Chem. 281: 40089-40095

Teriete P, Banerji S, Noble M, Blundell C, Wright A, Pickford A, Lowe E, Mahoney D, Tammi M, Kahmann J, Campbell I, Day A, Jackson D. (2004) Structure of the Regulatory Hyaluronan-Binding Domain in the Inflammatory Leukocyte Homing Receptor Cd44 Mol Cell. 13: 483-496.

Ulrich EL, Akutsu, 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. (2007) BioMagResBank. Nucleic Acids Research 36, D402-D408.

Vila JA, Ripoll DR, Baldoni HA, Scheraga HA (2002) Unblocked statistical-coil tetrapeptides and pentapeptides in aqueous solution: a theoretical study. J Biomol

NMR 24:245-262

Vila JA, Villegas ME, H.A. Baldoni and Scheraga HA. (2007) Predicting $^{13}C^{\alpha}$ chemical shifts for validation of protein structures. Journal of Biomolecular NMR, 38, 221-235.

Vila JA, H.A. Baldoni, Scheraga HA (2009) Performance of Density Functional Models to Reproduce Observed $^{13}C^{\alpha}$ Chemical Shifts of Proteins in Solution. Journal of Computational Chemistry, 30, 884-892.

Wishart DS, Bigam CG, Yao J, Abildgaard F, Dyson HJ, Oldfield E, Markley JL, Sykes BD (1995) ^{1}H, ^{13}C and ^{15}N chemical shift referencing in biomolecular NMR. J Biomol NMR 6:135-140

Wlodawer A, Walter J, Huber R, Sjolin L. (1984) Structure of bovine pancreatic trypsin inhibitor. Results of joint neutron and X-ray refinement of crystal form II J Mol Biol, 180: 301-329

Wlodawer A, Nachman J, Gilliland GL, Gallagher W, Woodward C. (1987) Structure of form III crystals of bovine pancreatic trypsin inhibitor. J Mol Biol, 198: 469-480

Wu B., Arumugam S, Gao G., Lee G.I., Semenchenko V., Huang W, Brew K, Van Doren S.R. (2000) NMR structure of tissue inhibitor of metalloproteinases-1 implicates localized induced fit in recognition of matrix metalloproteinases. J Mol Biol, 295: 257-268

Zahn, R., Damberger, F., Ortenzi, C., Luporini, P., Wuthrich, K. (2001) NMR Structure of the Euplotes Raikovi Pheromone Er- 23 and Identification of its Five Disulfide Bonds J Mol Biol, 313: 923-931.

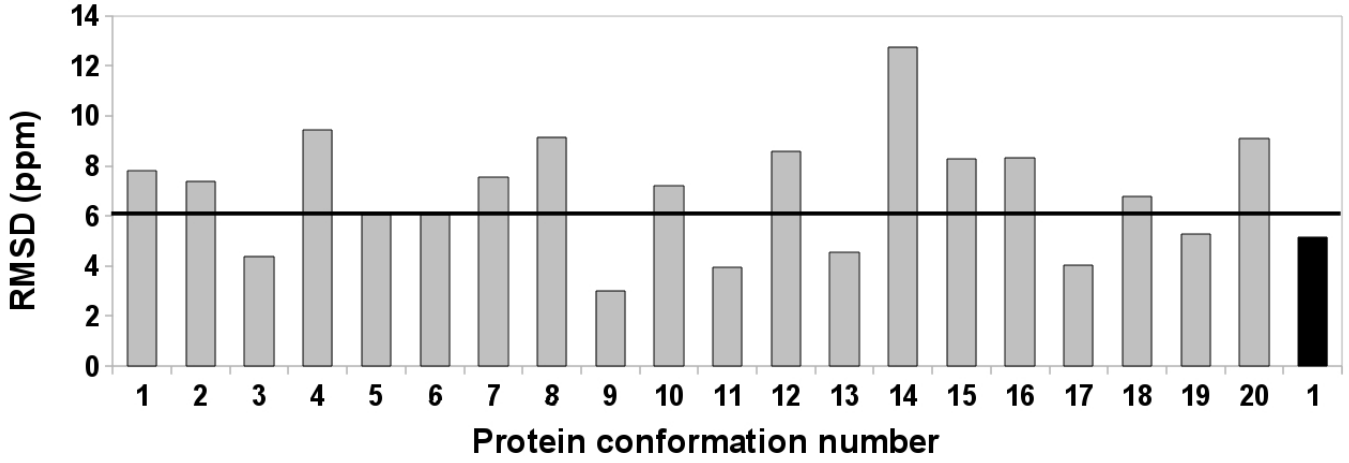
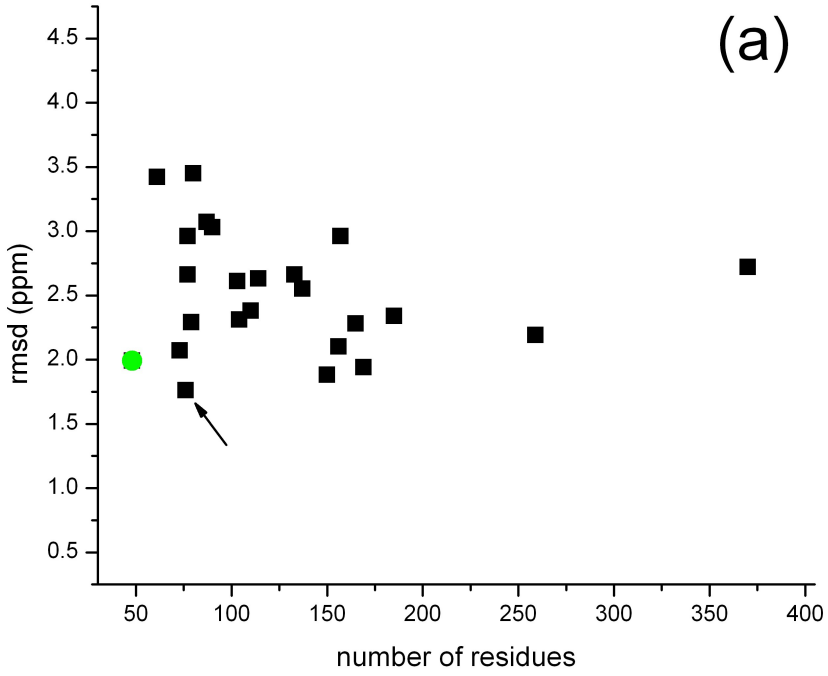


Figure S1. Grey filled bars indicate the rmsd (ppm), between the observed and computed $^{13}C^{\alpha}$ chemical shifts for the cysteines in the 2 cystines, for each of the 20 NMR-determined conformations of the MIA protein (1HJD). Black filled bar indicates the rmsd (ppm) computed for the X-ray determined structure of the MIA protein (1I1J). The solid horizontal line (6.08 ppm) indicates the *ca*-rmsd value computed from the 20 NMR-determined conformations of 1HJD.


(a)
rmsd (ppm)
number of residues
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
50
100
150
200
250
300
350
400

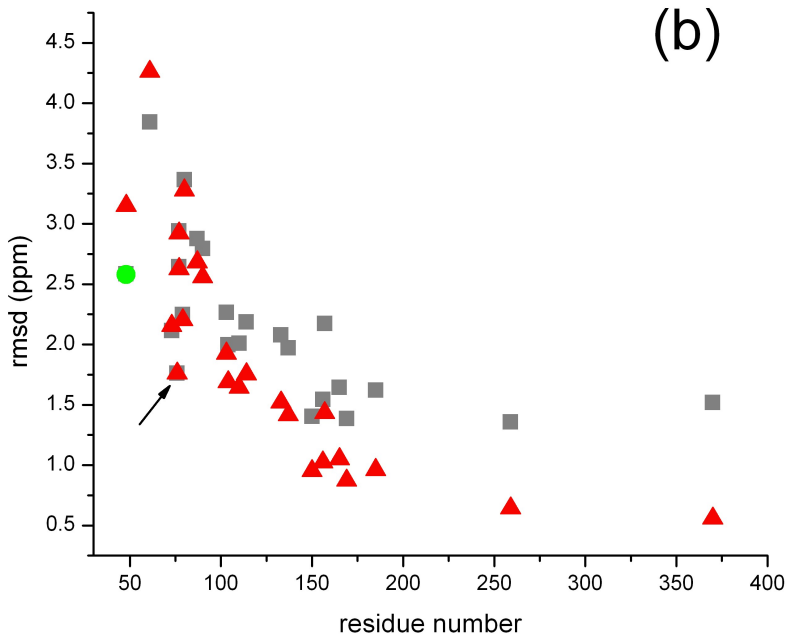

(b)
rmsd (ppm)
residue number
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
50
100
150
200
250
300
350
400

Figure S2. (**a**) rmsd between predicted $^{13}C^{\alpha}$ chemical shifts, by using the *Che*Shift server (Vila et al., 2009), and the observed values, for the first model (if more than one model exists) of the 24 NMR-determined structures listed above. Green-filled circle denotes the rmsd for the smallest (48 residues) structure of 2JVD. The arrow points to the rmsd of 1D3Z (ubiquitin) taken as a reference; (**b**) Grey-filled squares denote the $rmsd_{76}$, computed as described by Eq. (1) in the main text, and red-filled triangles the rmsd-per-residue for each the 24 structures mentioned in (**a**); it should be noted that *all* the rmsd-per-residue values (shown as red-filled triangles) have been shifted, for a better graphic representation, by a fixed amount, namely by multiplying each of them by 76 without affecting the relative values of the rmsd's; green-filled circle denotes the $rmsd_{76}$ for protein 2JVD, and the arrow points to the $rmsd_{76}$ of model 1, out of 10 models, of 1D3Z (ubiquitin) adopted here as a reference.

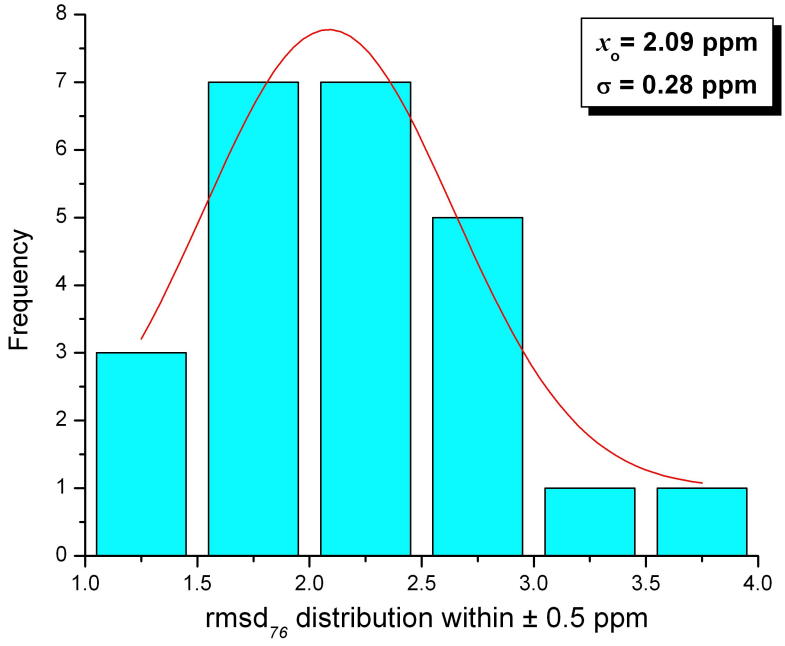


Figure S3. Bars represent the frequency of the rmsd$_{76}$ distribution, within ± 0.5 ppm, for 24 proteins whose rmsd$_{76}$ values are shown in Figure S2b, as grey-filled squares. The parameters for a Gaussian or Normal distribution (solid red line) that fits the data are inserted as a panel