J Biomol NMR (2012) 52:351–364
DOI 10.1007/s10858-012-9615-8

ARTICLE

Simultaneous single-structure and bundle representation of protein NMR structures in torsion angle space

Daniel Gottstein · Donata K. Kirchner · Peter Güntert

Received: 9 December 2011 / Accepted: 3 February 2012 / Published online: 22 February 2012

Abstract A method is introduced to represent an ensemble of conformers of a protein by a single structure in torsion angle space that lies closest to the averaged Cartesian coordinates while maintaining perfect covalent geometry and on average equal steric quality and an equally good fit to the experimental (e.g. NMR) data as the individual conformers of the ensemble. The single representative 'regmean structure' is obtained by simulated annealing in torsion angle space with the program CYANA using as input data the experimental restraints, restraints for the atom positions relative to the average Cartesian coordinates, and restraints for the torsion angles relative to the corresponding principal cluster average values of the ensemble. The method was applied to 11 proteins for which NMR structure ensembles are available, and compared to alternative, commonly used simple approaches for selecting a single representative structure, e.g. the structure from the ensemble that best fulfills the experimental and steric restraints, or the structure from the ensemble that has the lowest RMSD value to the average Cartesian coordinates. In all cases our method found a structure in torsion angle space that is significantly closer to the mean coordinates than the alternatives while maintaining the same quality as individual conformers. The method is thus suitable to generate representative single structure representations of protein structure ensembles in torsion angle space. Since in the case of NMR structure calculations with CYANA the single structure is calculated in the same way as the individual conformers except that weak positional and torsion angle restraints are added, we propose to represent new NMR structures by a 'regmean bundle' consisting of the single representative structure as the first conformer and all but one original individual conformers (the original conformer with the highest target function value is discarded in order to keep the number of conformers in the bundle constant). In this way, analyses that require a single structure can be carried out in the most meaningful way using the first model, while at the same time the additional information contained in the ensemble remains available.

Keywords Regularization · NMR structure · CYANA · REGMEAN · CYRANGE

D. Gottstein · D. K. Kirchner · P. Güntert (✉)
Institute of Biophysical Chemistry, Center for Biomolecular Magnetic Resonance, and Frankfurt Institute for Advanced Studies, Goethe University Frankfurt am Main, Max-von-Laue-Str. 9, 60438 Frankfurt am Main, Germany
e-mail: guentert@em.uni-frankfurt.de

P. Güntert
Graduate School of Science, Tokyo Metropolitan University, Hachioji, Tokyo, Japan

Introduction

Traditionally, the NMR solution structure of a protein is represented by a bundle of conformers that are calculated using identical input data but starting from different randomized initial conformations. Structure bundles are a means to represent multiple conformations that are compatible with the experimental data, for instance in the case of side-chains or loops, and they can be useful to convey an impression of how well the experimental data determines the 3D structure, although the spread of a structure bundle represents in practice only its precision rather than the accuracy (Zhao and Jardetzky 1994). On the other hand, it is often desirable to represent a protein structure by a single "representative conformer" as it is done, for instance, for X-ray crystal structures, and is required for many analyses of protein structures, or for their use in the modeling of

protein–ligand complexes in pharmaceutical research (Pellecchia et al. 2002).

Various approaches to obtain a single representative structure from an ensemble of structures have been used and compared (Dukka 2009; Nilges et al. 1988; Sutcliffe 1993; Thomas and Pastore 2005). One of the ensemble members can be selected as the representative structure, for example the "best conformer" that violates least the experimental restraints or has the lowest energy, or the "conformer closest to the mean" that has the smallest RMSD to the mean coordinates obtained by averaging the Cartesian coordinates after superimposing the ensemble members for minimal RMSD. The mean coordinates themselves, although by definition representing the center of a bundle of structures in conformation space, are not suitable because the averaging results in deviations of the covalent structure from optimal values and steric clashes. Alternatively, a new "regularized mean structure" can be calculated that should be close to the mean coordinates but has ideal covalent geometry and is in good agreement with the experimental data and stereochemical requirements. This approach has been implemented in Cartesian space by calculating the mean structure of the ensemble followed by restrained energy refinement to rectify the structure (Nilges et al. 1988). Because the regularization takes place in Cartesian space, care has to be taken to avoid that no unacceptably large residual distortions of the covalent structure remain after the restrained energy refinement. This problem is compounded by the fact that most protein structures contain unstructured regions for which coordinate averaging results in particularly large deviations from ideal polypeptide geometry.

Here we present a new method that works in torsion angle space and thus preserves the covalent geometry exactly. Our approach is very similar to a standard CYANA structure calculation (Güntert et al. 1997). It is carried out starting from random initial structures rather than from the averaged coordinates, and uses the same experimental restraints as for the original structure calculation. The experimental restraints are supplemented by weak restraints that guide the structure as close towards the mean structure of the bundle as it is compatible with the rigidly maintained covalent structure. The method does not aim at improving the precision or accuracy of a structure.

Materials and methods

Regularized mean structure

To compute the regularized mean structures in torsion angle space, we implemented a new algorithm, REGMEAN, into the CYANA software package for NMR structure determination (Güntert 2009; Güntert et al. 1997). The REGMEAN algorithm (Fig. 1) performs a structure calculation using simulated annealing by molecular dynamics simulation in torsion angle space. The original protein sequence is extended by a sterically invisible, flexible linker and an artificial pseudo-residue that is internally rigid and comprises all atoms of the reference structure to which the regularized mean structure should be as close as possible. The REGMEAN algorithm automatically appends an entry for the artificial pseudo-residue to the standard CYANA residue library. In contrast to the library entries for normal amino acid residues that contain torsion angle definitions for the internal degrees of freedom of an amino acid residue, the artificial pseudo-residue is rigid and has no rotatable torsion angles. It thus maintains the reference conformation exactly throughout the structure calculation.

For the purpose of computing a regularized mean structure, the reference conformation consists of the mean coordinates of the backbone atoms N, C^{α}, and C' of the ordered residues that can be well superimposed globally, as determined by the CYRANGE algorithm for the objective identification of residue ranges for the superposition of protein structures (Kirchner and Güntert 2011). Alternatively, the subset of protein atoms for the calculation of the mean coordinates can be chosen explicitly. The individual conformers of the input structure bundle are superimposed on the first one for minimal backbone RMSD of the ordered residues, and the mean coordinates are obtained as the arithmetic mean over the bundle of superimposed conformers. Note that for other applications of this torsion angle space regularization algorithm, the reference structure may be chosen differently. For instance, to adapt a single protein structure that was originally computed with other software, e.g. an X-ray crystal structure, to the standard geometry of CYANA, the reference structure would be the entire original structure rather than the mean backbone coordinates of a structure bundle.

The input data for the calculation of the regularized mean structure is then augmented by upper bounds of 0.01 Å between all corresponding atoms in the protein and in the reference structure. These additional distance restraints are applied with a minute weight of 0.01 relative to the weight of the experimental NOE distance restraints.

In addition, torsion angle restraints are generated for all torsion angles that are defined by non-hydrogen atoms for which coordinates are available in the input structure bundle. To this end, the average value $\overline{\phi}$ and the standard deviation σ_ϕ of a torsion angle are calculated taking into account the periodicity:

$$\overline{\phi} = \arg \sum_j e^{i\phi_j} \tag{1}$$

$$\sigma_\phi = \sqrt{\sum_j \min\left(|\phi_j - \overline{\phi}|, 2\pi - |\phi_j - \overline{\phi}|\right)^2}. \tag{2}$$

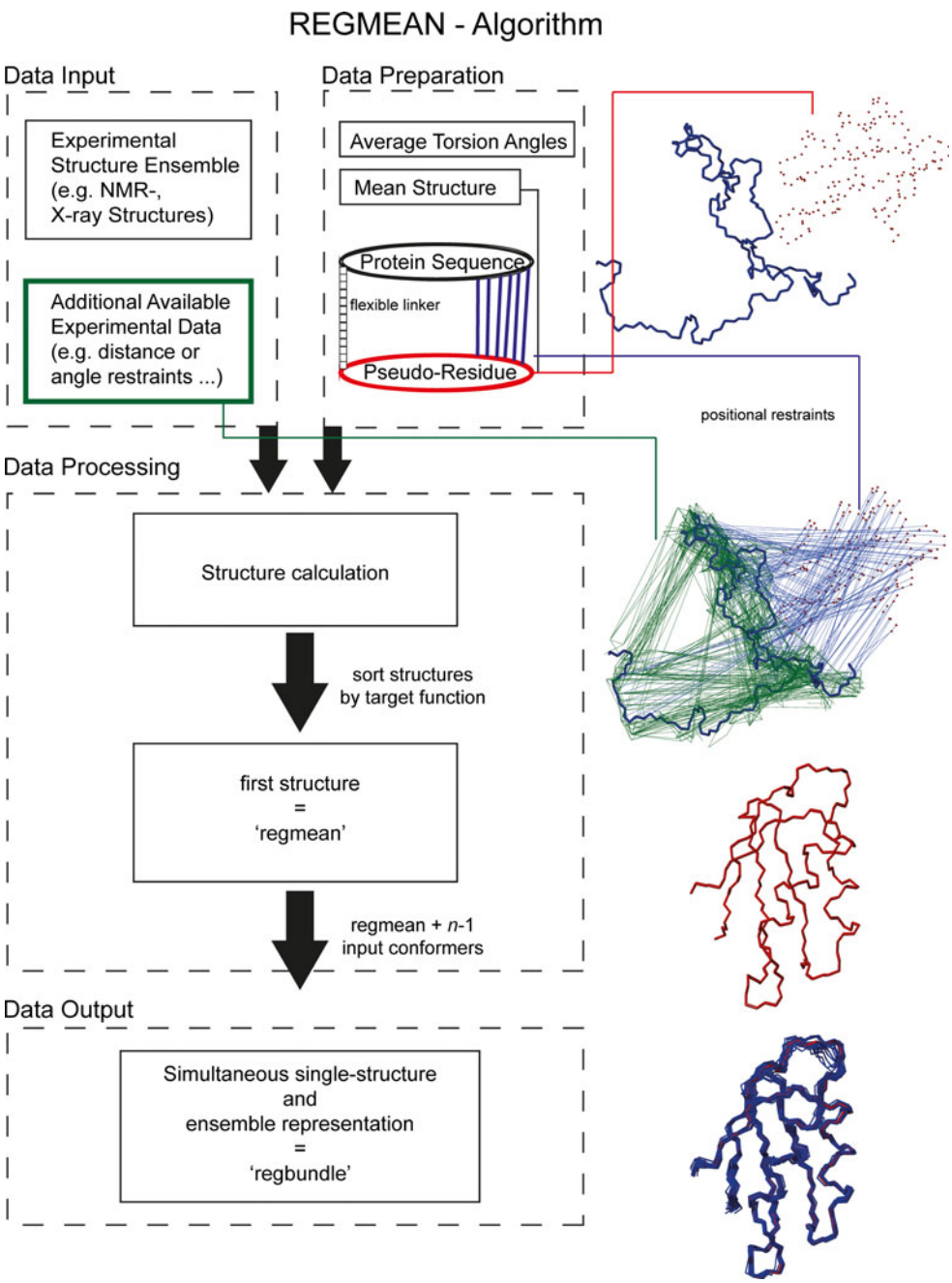

Fig. 1 Flowchart of the REGMEAN algorithm

The sums run over all conformers in the input structure bundle, and arg denotes the argument of a complex number, i.e. $\arg z = \phi \,(\mathrm{mod}\, 2\pi)$ for a complex number $z = a\,e^{i\phi}$. Equations (1) and (2) give meaningful results if the torsion angle has a unimodal distribution but can be misleading for a torsion angle whose values fall into multiple, distinct clusters, as it occurs frequently for side-chain torsion angles. Therefore, also bimodal average values $\overline{\phi}_1$ (for the more populated cluster) and $\overline{\phi}_2$ and the corresponding standard deviation $\sigma_\phi^{(2)}$ are calculated as described in the Appendix. If $\sigma_\phi \leq 2.5\sigma_\phi^{(2)}$ or if the largest cluster of torsion angle values comprises less than half of the conformers, then a unimodal distribution is assumed and the torsion angle is restrained to the range $\overline{\phi}_1 \pm \max(1.5\sigma_\phi, 10^\circ)$. Otherwise, the more populated cluster of the bimodal distribution is used to restrain the torsion angle to the range $\overline{\phi}_1 \pm \max(1.5\sigma_\phi^{(2)}, 10^\circ)$. These torsion angle restraints are applied with a small weight of 0.02 relative to the standard weight for experimental torsion angle restraints.

The parameters 2.5, 1.5, and 10° are default values that have been used for all calculations of this paper, but can be changed if necessary. Small (±25%) variations of these parameters did not lead to significant changes of the results. The threshold of 2.5 in the condition $\sigma_\phi \leq 2.5\sigma_\phi^{(2)}$ for discriminating between unimodal and bimodal distributions was chosen such that a (narrower) torsion angle restraint based on the more populated cluster of the bimodal distribution is chosen only if the bimodality is very clear. In all other cases, a more conservative (wider) restraint is based on the distribution of all torsion angle values in the bundle. The scaling factor 1.5 for the width of the restraint relative to the standard deviation of the torsion angle values was chosen such that on average the allowed ranges of the restraints include more than 85% of the torsion angle values in the input conformers. A minimal half-width of 10° was chosen to avoid extremely narrow torsion angle restraints in the case of locally tight input structure bundles.

With the additional distance and torsion angle restraints that are designed to guide the structure towards reference coordinates, a structure calculation is performed using the same experimental input data and parameters (e.g. number of torsion angle dynamics steps) as the structure calculation that yielded the input structure bundle. The structure calculation is started from 25 conformers with random torsion angle values. Since the additional restraints have a very low weight, they interfere only marginally with the original experimental restraints, and in general it is possible to find solutions that fulfill the latter as well as in the conventional structure calculation without additional restraints. Rather, the additional restraints serve to select among all (almost) equally good solutions the one that is closest to the mean coordinates but in perfect agreement with the standard geometry. The conformer with the lowest final target function is chosen as the regularized mean structure determined by the REGMEAN algorithm, to which we refer as the '*regmean structure*'.

This conformer forms, together with the $n - 1$ conformers with lowest target function values among the $n = 20$ conformers of the input structure bundle, a new bundle consisting of a single representative structure as the first conformer and all but one of the original individual conformers. This '*regmean structure bundle*' or '*regmean bundle*' provides a simultaneous single-structure and ensemble representation of protein NMR structures in torsion angle space.

Multi-domain proteins

The aforementioned procedure to produce a regularized mean structure in torsion angle space needs to be modified for proteins with multiple domains whose relative orientation is less well defined than the structures of the individual domains. In this case, the CYRANGE algorithm yields separate well-defined residue ranges for the individual domains. For each of these domains, a separate mean structure is computed by superimposing the backbone coordinates of the corresponding ordered residues. Accordingly, the REGMEAN algorithm extends the protein sequence by multiple sterically invisible, flexible linkers and artificial pseudo-residues, one for each domain, and adds an entry for each artificial pseudo-residue to the standard CYANA residue library. As a result of minimizing the deviation from all domain-specific mean structures simultaneously, the REGMEAN algorithm still produces a single regularized mean structure, which is in optimal agreement with the mean coordinates of all domains.

Multimeric proteins and complexes

The REGMEAN algorithm is also applicable to multi-component systems, for example symmetric dimers, and protein–protein complexes. These can be treated in the same way as monomeric proteins with one or several domains.

Alternative single structure representations of a structure bundle

Three alternatives to our torsion angle space regularized mean structure were considered. (1) The unregularized mean coordinates ('mean') obtained by computing the arithmetic mean of the coordinates of the individual conformers of the structure bundle after optimally superimposing the individual conformers onto the first conformer for lowest RMSD values of the backbone atoms N, C^α, C' in the ordered region of the protein. This structure lies optimally in the center of the structure bundle but distortions of the covalent geometry and violations of conformational and steric restraints resulting from the coordinate averaging make it unsuitable to represent the NMR solution structure of a protein. (2) The individual conformer from the structure bundle that has the best CYANA target function value ('best conformer'). (3) The individual conformer from the structure bundle that is closest to the mean ('closest conformer'). This structure generally fulfills the conformational restraints similarly well as other conformers from the bundle, and is located closest to the mean among the individual conformers of the structure bundle. This choice of a representative structure has the advantage that it is defined by the coordinates of the structure bundle alone. This structure fulfills best the conformational restraints among the conformers of the structure bundle but is not necessarily located near the center of the structure

bundle. Note that the mean coordinates (1) and the closest conformer have the additional disadvantage that for multi-domain proteins they do not define a single representative conformer but in general different ones for each domain, whereas the best conformer and the regmean structure are both single structures that represent sensibly the complete multi-domain protein.

NMR data sets

Single structure representations were assessed for NMR structure bundles of 11 proteins for which the NMR solution structure had been determined earlier, and which are referred to by four-letter acronyms (Table 1). The proteins copz (Wimmer et al. 1999), cprp (Calzolai et al. 2005), enth (López-Méndez and Güntert 2006; López-Méndez et al. 2004), fsh2 (Scott et al. 2004, 2005), pbpa (Horst et al. 2001), rhod (Pantoja-Uceda et al. 2005; Pantoja-Uceda et al. 2004), and wmkt (Antuch et al. 1996) are proteins with a well-defined single-domain structure. The proteins cprp, enth, pbpa, and scam are predominantly α-helical; wmkt and ww2d are pure β-sheet proteins; copz, fsh2, rhod and smbp have mixed α/β secondary structure. The protein fspo has an unusual, less well-defined fold with little regular secondary structure (Pääkkönen et al. 2006). The proteins scam and smbp are proteins with two domains connected by a flexible linker, and smbp is one of the largest proteins whose detailed structure has been solved by NMR (Kainosho et al. 2006). The NMR data sets for scam and smbp were obtained with the help of stereo-array isotope labeling (SAIL). The protein ww2d forms a symmetric dimer (Ohnishi et al. 2007). The structures and the corresponding restraints are available for download from http://www.cyana.org/regmeandata.tgz.

Two structure bundles were considered for each of these proteins: the final structure bundle, and the structure bundle obtained in the initial cycle 1 of automated NOE assignment and structure calculation (Herrmann et al. 2002) with CYANA (Güntert 2003), i.e. all structures were recalculated using the experimental chemical shift lists, NOESY peak lists, and (for cprp, pbpa, smbp, and ww2d) torsion angle restraints. This enabled comparisons of the REGMEAN results for two structure bundles of different precision and quality for each of the proteins.

For comparison, we also applied the REGMEAN algorithm to seven NMR data sets for the Northeast Structural Genomics Consortium target proteins HR5460A (PDB code 2LAH), HR6430A (PDB code 2LA6), HR6470A (PDB code 2L9R), and OR36 (PDB code 2LCI) that were provided for the Critical Assessment of Structure Determination by NMR (CASD-NMR) project (Rosato et al. 2009, 2012). For all four proteins the refined final data set was available. In addition, unrefined, "raw" data sets were available for the proteins HR6430A, HR6470A, and OR36. All data sets can be downloaded from http://www.wenmr.eu/wenmr/casd-nmr-data-sets.

The recalculation of structures with the standard CYANA algorithm (Güntert 2009) used the original NOESY peak lists for seven cycles of automated NOESY assignment (Herrmann et al. 2002) and structure calculation by torsion angle dynamics (Güntert et al. 1997), followed by a

Table 1 Summary of proteins and NMR data for which regularized mean structures were computed

Acronym	PDB code	Residues	Distance restraints	Angle restraints	Ordered residues	Helix (%)	Sheet (%)
copz	1CPZ	68	1,013	–	3–7, 15–68	25	22
cprp	1U3 M	117	1,966	123	135–145, 155–200, 213–241	50	2
enth	1VDY	140	3,361	–	9–130	63	0
fsh2	1WQU	114	2,272	–	8–110	23	25
fspo	1VEX	56	796	–	614–666	0	14
pbpa	1GM0	142	2,150	148	9–142	63	0
rhod	1VEE	134	2,996	–	6–82, 86–125	37	10
scam	1X02	148	2,403	–	5–72; 85–145	66	0
smbp	2D21	370	3,768	462	126–224, 248–256, 331–370; 4–109, 261–309	43	6
wmkt	1WKT	88	1,368	–	3–88	0	40
ww2d	2DWV	2 × 49	1,611	44	2 × 14–43	0	22

Ordered residues were determined by the CYRANGE algorithm (Kirchner and Güntert 2011). In the case of proteins with multiple domains (scam and smbp) the ranges for the individual domains are separated by semicolons. Percentages of residues in helices and β-sheets were obtained from the HELIX and SHEET secondary structure records in the original PDB files given in the second column. Publications of the original NMR structure determinations: copz (Wimmer et al. 1999), cprp (Calzolai et al. 2005), enth (López-Méndez and Güntert 2006; López-Méndez et al. 2004), fsh2 (Scott et al. 2004, 2005), fspo (Pääkkönen et al. 2006), pbpa (Horst et al. 2001), rhod (Pantoja-Uceda et al. 2004, 2005), scam and smbp (Kainosho et al. 2006), wmkt (Antuch et al. 1996), ww2d (Ohnishi et al. 2007)

final structure calculation using only unambiguously assigned NOE distance restraints. Structure calculations were started from 100 (200 for the largest protein smbp) conformers with random torsion angle values, 10,000 (15,000 for scam, 20,000 for smbp and ww2d) torsion angle dynamics steps were applied, and from the resulting bundles the 20 conformers with the lowest final target function values were selected to form the NMR structure bundle that represents the solution structure of the protein and that was used as the input structure bundle for computing the regularized mean structure.

Structure analysis

CYANA was used to obtain statistics on target function values and restraint violations, and to compute RMSD values. Target function values were computed for the experimental and steric restraints. RMSD values for a structure bundle were computed to the mean coordinates after superposition of the backbone atoms N, C^{α}, and C' of the structured regions of the proteins, treating each domain separately. For a bundle with n conformers the RMSD reported is the average of the n RMSDs of the individual conformers to the mean coordinates (not the average of the $n(n-1)/2$ pairwise values). The RMSD between two bundles is the RMSD between their mean coordinates. The program MOLMOL (Koradi et al. 1996) was used to visualize 3D structures.

Structure validation

Various structure validation scores were evaluated with the help of a new CYANA script, 'validate', that calls the respective structure validation programs and collects their results in a concise validation report. The following validation parameters were computed for each conformer of the original structure bundles and for the regmean structures: (1) One of the scores (zp-comb) calculated by ProSa2003 (Sippl 1993). (2) The Verify3D score (Bowie et al. 1991; Lüthy et al. 1992). (3) The clashscore calculated by MolProbity, and the MolProbity score, which takes into account steric clashes, and Ramachandran plot and staggered rotamer outliers (Chen et al. 2010; Davis et al. 2004, 2007). (4) The packing, the Ramachandran plot appearance, the χ^1/χ^2 rotamer normality, and the backbone conformation quality scores of the WHAT_CHECK program (Hooft et al. 1996). (5) The percentage of residues in the most favored region of the Ramachandran plot (Rama G-factor), and the χ^1 rotamer normality (Chi-1 G-factor), as defined by the program PROCHECK-NMR (Laskowski et al. 1996; Morris et al. 1992). (6) The LGscore and the MaxSub score of the ProQ program (Wallner and Elofsson 2003).

Results and discussion

To evaluate the suitability of the REGMEAN algorithm to provide representations of NMR solution structures, we computed the regmean structure and the regmean structure bundle, as well as the abovementioned alternative single structure representations for 11 proteins, and compared them with the corresponding original structure bundles (Fig. 2). The ideal regularized mean structure would have a target function value equal to or below the one of the best input conformer, and would coincide perfectly with the mean coordinates of the input structure bundle. The latter aim cannot be met perfectly because the averaging introduces into the mean coordinates deviations from ideal values of bond lengths, bond angles, and planar groups, as well as steric clashes. Some deviation of the regularized mean structure from the mean coordinates must occur to avoid these unrealistic situations. Here we show that the REGMEAN algorithm produces regularized mean structures that fulfill the two simultaneous criteria of low target function values and small RMSDs from the mean coordinates much better than alternative single structure representations.

Table 2 shows the target function values of the original structure (from the final structure calculation of the recalculation with automated NOESY assignment), mean coordinates, best (lowest target function) original conformer, the closest-to-mean original conformer, the (single) regmean structure, and the regmean bundle for the 11 proteins. In all cases the target function value of the regmean structure is among that of the best 8 (of all 20) conformers of the input structure bundle ('rank' column in Table 2). For seven of the 11 proteins the target function value of the regmean structure is as good as or better than for the best or second best input conformers. Relative to the average target function values of the conformers of the input conformers, the regmean target function values vary between 7 and 106% with an average of 68%, but they never exceed the maximal target function value of the input structure bundle. Note that a relative target function value above 100%, as in the case of pbpa, does not imply that the regmean is worse than the original structure, since there are still several input conformers with higher target function values than the regmean structure. This shows that the application of the additional restraints that direct the structure towards the mean coordinates and the most populated torsion angle ranges do not increase the violations of the experimental restraints. Thus, the regmean structure fulfills the experimental restraints equally well as the majority of the individual conformers of the input structure bundle. In contrast, the distortions inherent in the mean coordinates lead to unrealistically high target function values between 93 and 4.8×10^5 Å^2. The particularly high target function values for the mean coordinates of the

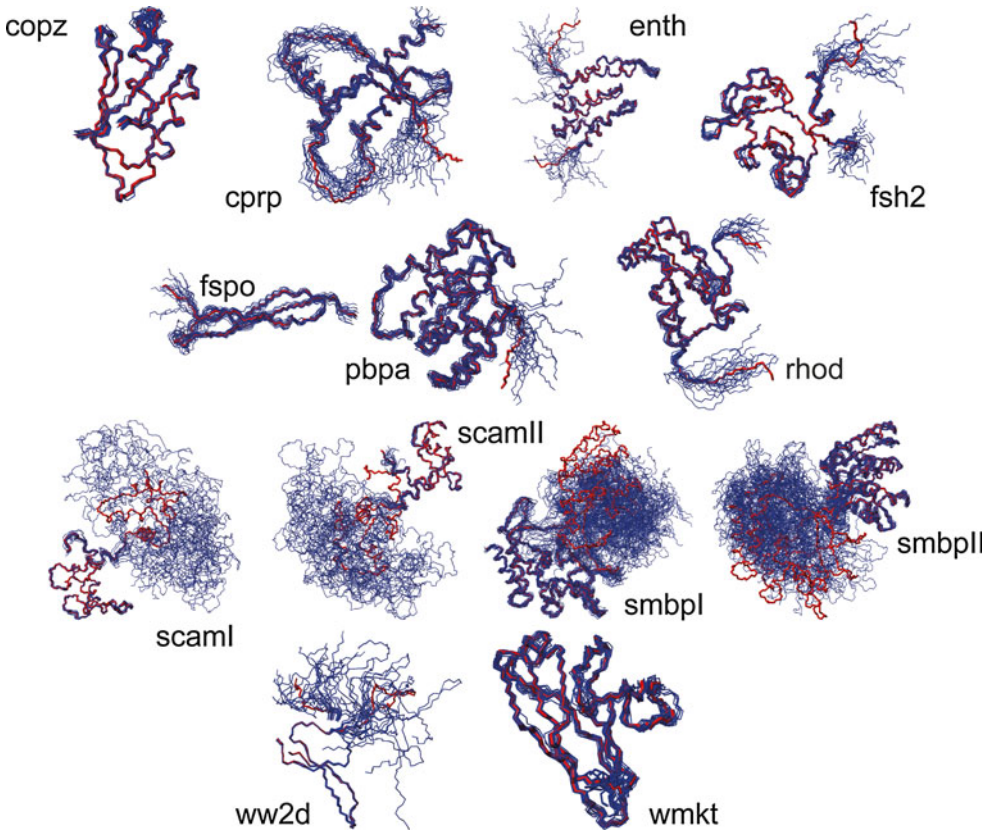

Fig. 2 Simultaneous single-structure and bundle representation of structure ensembles in torsion angle space for 11 proteins (see "Materials and methods"). Proteins are identified by their acronyms in Table 1. Individual conformers are shown in *blue*, the regularized mean structure in *red*. Two separate superpositions are shown for the two-domain proteins scam and smbp

Table 2 CYANA target function values (Å^2)

Protein	Reference bundle	Mean	Best	Closest	Regmean	Regmean bundle	Rank
copz	0.02 (0.02–0.03)	93	0.02	0.02	0.02 (88%)	0.02 (0.02–0.03)	7
cprp	1.95 (1.40–2.31)	3,575	1.40	1.56	1.42 (73%)	1.91 (1.40–2.31)	2
enth	0.32 (0.15–0.49)	5,793	0.15	0.15	0.12 (38%)	0.31 (0.12–0.44)	1
fsh2	1.41 (1.20–1.50)	375	1.20	1.20	1.08 (77%)	1.39 (1.08–1.50)	1
fspo	0.27 (0.05–0.54)	179	0.05	0.06	0.05 (19%)	0.25 (0.05–0.51)	2
pbpa	0.16 (0.05–0.22)	4,066	0.05	0.17	0.17 (106%)	0.16 (0.05–0.22)	8
rhod	0.76 (0.53–0.91)	854	0.53	0.56	0.53 (70%)	0.74 (0.53–0.91)	1
scam	0.46 (0.45–0.47)	3.6×10^5; 4.8×10^5	0.45	0.46; 0.48	0.46 (100%)	0.47 (0.45–0.48)	7
smbp	3.61 (3.06–3.95)	7,811; 9,116	3.07	3.10; 3.56	0.46 (100%)	0.47 (0.45–0.48)	7
wmkt	0.11 (0.01–0.20)	275	0.01	0.12	0.01 (7%)	0.10 (0.01–0.19)	2
ww2d	0.25 (0.08–0.40)	2,921	0.079	0.35	0.10 (39%)	0.24 (0.08–0.38)	2

Target function values were computed including all experimental and steric distance and torsion angle restraints given in Table 1 for the following structures: 'Reference bundle', original structure bundle (from the final structure calculation of the recalculation with automated NOESY assignment); 'Mean', mean coordinates; 'Best', original conformer with the lowest target function value; 'Closest', original conformer that is closest to the mean coordinates; 'Regmean', regmean conformer; 'Regmean bundle', regmean structure bundle consisting of the regmean conformer and the 19 original conformers with lowest target function values. For structure bundles the average and, in parentheses, the minimum and the maximum value of the target function are given. For the two-domain proteins scam and smbp two values are given for the 'Mean' and 'Closest' structures, corresponding to the two domains given in Table 1. The percentage number given for the regmean structure denotes the target function value relative to the average target function value of the reference structure bundle. The last column, 'Rank', indicates the position at which the regmean structure was ranked among the reference conformers upon ordering according to increasing target function values. For instance, rank 1 means that the regmean structure had a better (lower) target function value than the best conformer of the reference structure

two-domain protein scam result from the severe artifacts that are introduced by averaging the coordinates of domain 2 after superposition of domain 1, and vice versa. The great variation in the position of the domain that is not superimposed leads to a collapsed, overly compact set of its mean coordinates with a very large number of steric clashes. In more than 80% of the cases the regmean structure has also a lower target function value than the input conformer that is closest to the mean coordinates.

RMSD values for these structures in Table 3 show that for all proteins the REGMEAN algorithm found a regularized mean structure very close to the mean coordinates of the original structure bundle (Fig. 2). The RMSD values between the (single) regmean structure and the mean coordinates of the original structure bundle varied between a minimum of 0.04 Å for copz and a maximum of 0.37 Å for fspo. These values are much smaller than the corresponding average RMSD values of the original structure bundles of 0.31 Å for copz and 1.20 Å for fspo. Over all 11 proteins the regmean RMSD values are between 7 and 33% with an average of 23% of the corresponding RMSD values of the input structure bundle. This proves that in all cases there exists a structure that is very close to the mean coordinates of the original structure bundle while maintaining correct geometry and agreement with the experimental data. For all proteins, the regmean structure fulfills these conditions much better than the alternative choices of the lowest target function conformer of the original structure bundle (column 'Best' in Table 3) or the individual conformer closest to the mean coordinates (column 'Closest' in Table 3). The RMSDs to the mean coordinates for the closest conformer are 1.7–10 times (average 3.5 times) larger than for the regmean structure, and in general even higher for the best conformer.

The regmean structure also lies at the center of the side-chain conformations of the original structure bundle, or of the principal cluster for a side-chain that adopts multiple distinct groups of conformations in the input bundle, as shown for one of the proteins (fsh2) in Fig. 3.

A comparison with respect to common structure validation scores between the regmean structures and the input structures is shown in Fig. 4. It shows that in the large majority of the cases the regmean structure lies roughly centrally within the range of values obtained for the conformers of the input structure bundles, and at times it has even obtained a better score. The number of cases in which the regmean structure is located at the bottom end of the range, or even outside, at an unfavorable score value (e.g. the Molprobity score for pbpa), is small. Furthermore, there appears to be no particular aspect of the regmean structure, measurable by one of the calculated structure validation scores, which differs from its counterpart in the original structure bundle in all of the examined proteins. Consequently, the regularization procedure described here generally yields structures which share the main characteristics of their parent structure bundles.

Table 3 RMSD values (Å)

Protein	Reference bundle	Mean	Best	Closest	Regmean	Regmean bundle	Regmean bundle to reference
copz	0.31 (0.21–0.42)	0.00	0.27	0.21	0.04 (13%)	0.29 (0.04–0.42)	0.02
cprp	1.04 (0.80–1.58)	0.00	0.89	0.80	0.27 (26%)	0.99 (0.26–1.56)	0.06
enth	0.50 (0.35–0.78)	0.00	0.35	0.35	0.10 (20%)	0.47 (0.10–0.78)	0.03
fsh2	0.45 (0.36–0.58)	0.00	0.36	0.36	0.15 (33%)	0.43 (0.14–0.59)	0.03
fspo	1.20 (0.64–1.58)	0.00	1.43	0.63	0.37 (31%)	1.14 (0.35–1.60)	0.07
pbpa	0.67 (0.50–0.88)	0.00	0.67	0.51	0.11 (16%)	0.65 (0.11–0.86)	0.03
rhod	0.41 (0.27–0.55)	0.00	0.31	0.27	0.11 (27%)	0.40 (0.11–0.55)	0.02
scam	0.31 (0.22–0.49)	0.00	0.31	0.22	0.08 (26%)	0.30 (0.07–0.49)	0.02
	0.30 (0.20–0.41)	0.00	0.40	0.20	0.02 (7%)	0.29 (0.03–0.41)	0.01
smbp	0.84 (0.56–1.20)	0.00	0.90	0.57	0.22 (26%)	0.81 (0.21–1.17)	0.04
	0.54 (0.39–0.76)	0.00	0.41	0.40	0.17 (31%)	0.53 (0.17–0.76)	0.02
wmkt	0.59 (0.37–0.83)	0.00	0.56	0.37	0.13 (22%)	0.56 (0.13–0.83)	0.03
ww2d	0.27 (0.18–0.37)	0.00	0.22	0.18	0.07 (26%)	0.26 (0.06–0.37)	0.02

RMSD values were computed for the backbone atoms N, C^{α}, and C′ of the ordered residues given in Table 1. For the two-domain proteins scam and smbp the two lines correspond to the two domains. For the 'Reference bundle' and the 'Regmean bundle' the average and, in parentheses, the minimum and the maximum value of the RMSD of the 20 individual conformers to the corresponding mean coordinates are given. For the 'Mean', 'Best', 'Closest', and 'Regmean' structures the RMSD to the mean coordinates of the reference bundle is given. The last column, 'Regmean bundle to reference', gives the RMSD between the mean coordinates of the reference bundle and the mean coordinates of the regmean bundle. The percentage number given for the regmean structure denotes the RMSD value relative to the average RMSD value of the reference structure

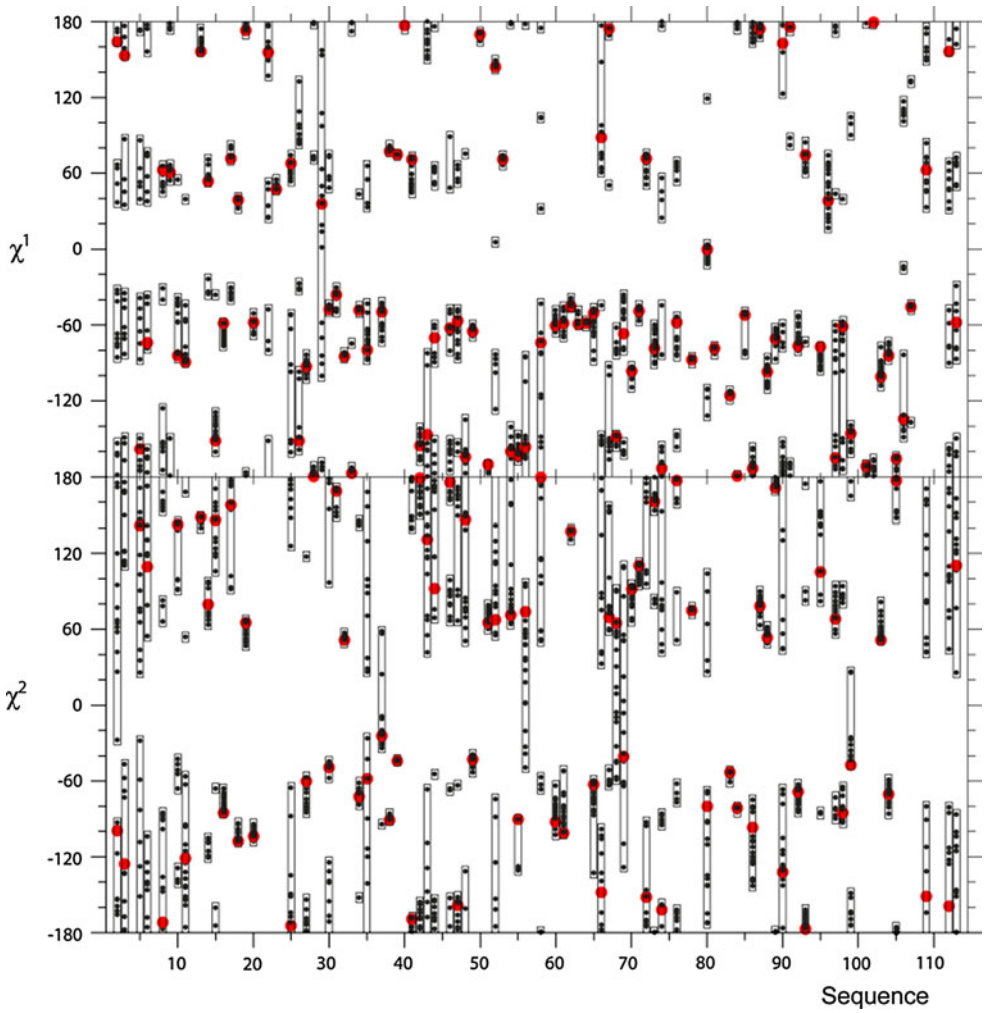

Fig. 3 Illustration of REGMEAN results for the χ^1 and χ^2 side chain torsion angles of the protein fsh2. *Red dots* represent torsion angle value of the regmean structure, *black dots* those of the 20 individual conformers of the original structure bundle

The regmean structure bundles formed by the single regmean structure and the 19 lowest target function input conformers have RMSD values to their mean coordinates that are very similar to those of the input structure bundle, and the RMSD shifts between the mean coordinates of the two structure bundles is only 0.01–0.07 Å (Table 3). The same holds true for the target function values (Table 2). The regmean structure bundle is thus an equally valid representation of the solution structure as the original structure bundle, and has the additional advantage to provide as its first conformer a single structure representation of the protein structure that lies as much as possible at the center of the bundle.

A similar situation as shown in Tables 2 and 3 for structures based on final structure bundles was observed also for structures based on the structure bundles from the initial cycle 1 of automated NOESY assignment with CYANA. The cycle 1 bundles are less well-defined and less in agreement with the experimental data than the final structure bundles, and thus present potentially more difficult targets for the REGMEAN procedure. They exhibited average target function values between 1.98 $Å^2$ for pbpa and 296 $Å^2$ for smbp, and average RMSDs to the mean coordinates between 0.56 Å for ww2d and 3.47 Å for domain 1 of scam. Nevertheless, the REGMEAN algorithm worked well also for the cycle 1 structures, for which it yielded regmean structures with, on average, target function values of 74% (range 49–94%) and RMSD values of 49% (range 23–80%) of the original average values. Remarkably, the target function value of the regmean structure was in all cases better than that of the best original cycle 1 conformer.

For an independent comparison, we also applied the REGMEAN algorithm to seven structures from the CASD-NMR data sets (see "Materials and methods"), which had never been considered during the development of the algorithm. The input structure bundles had average target function values of 0.27–15.86 $Å^2$, and average backbone RMSDs to the mean coordinates of 0.06–0.31 Å for the ordered regions. The REGMEAN algorithm yielded regmean structures with target function values of 87–99% and RMSD values of 11–60% of the original average values. In

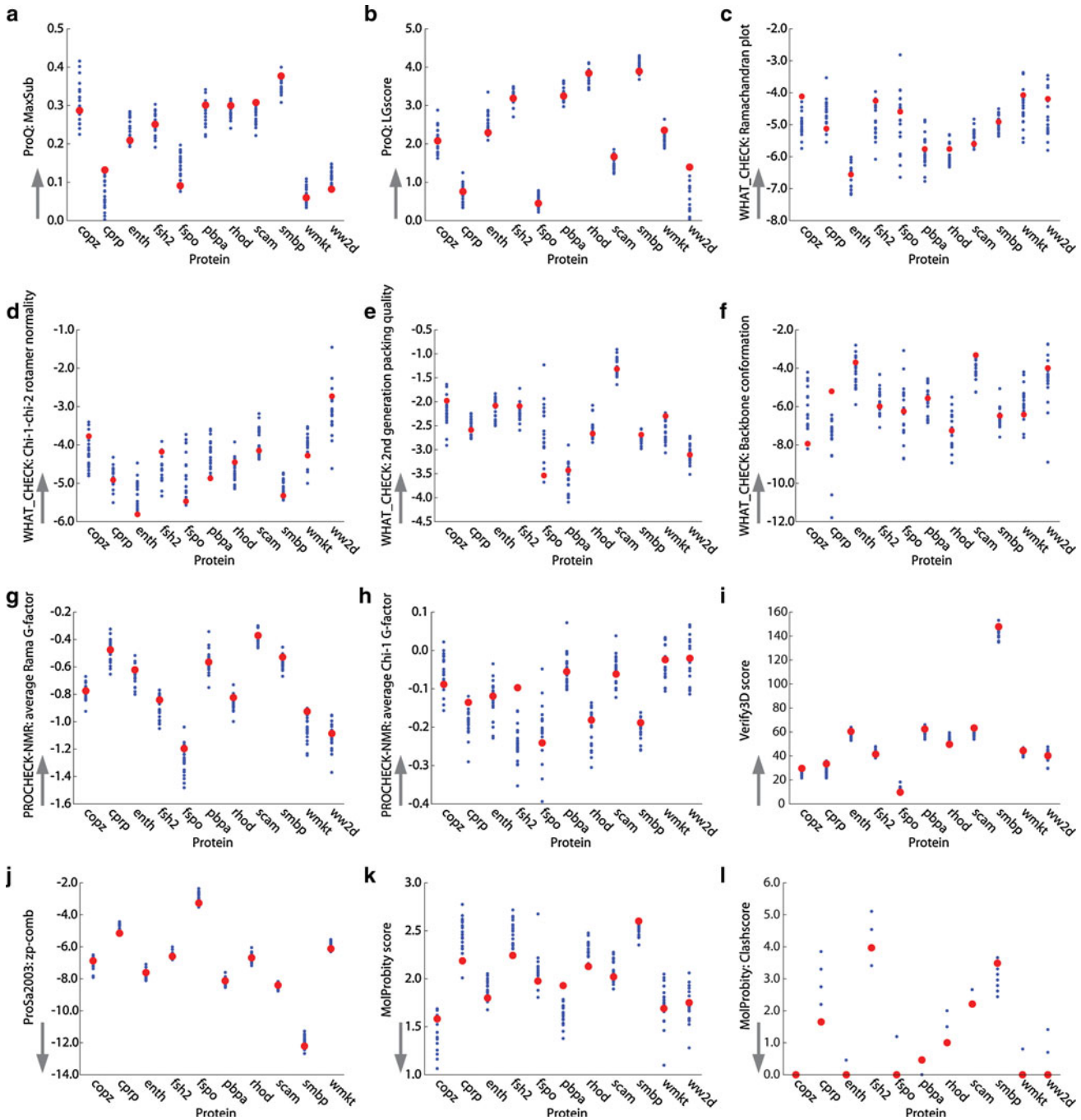

Fig. 4 Structure validation scores for the regmean structures (*red*) and the conformers of the input structure bundles (*blue*). The *arrow* to the *left* of each subfigure indicates the direction in which the score value becomes more favorable. **a**, **b** ProQ: MaxSub and LGscore. **c**–**f** WHAT_CHECK: Ramachandran plot appearance, χ^1/χ^2 rotamer normality, packing, and backbone conformation quality. **g**, **h** PROCHECK-NMR: percentage of residues in the most favored region of the Ramachandran plot (Rama G-factor), and the χ^1 rotamer normality (Chi-1 G-factor). Each point in the figure represents the value obtained when this particular score is averaged over all of the conformer's residues for which the score could be calculated. **i** Verify3D score. **j** ProSa2003 zp-comb score. When performing the calculations for the protein ww2d the ProSa2003 program crashed, thus no results can be displayed in this case. **k**, **l** MolProbity: MolProbity score and clashscore (number of serious clashes per 1,000 atoms)

all cases the target function value of the regmean structure was among that of the best 3 (of all 20) conformers of the input structure bundle. The algorithm thus yielded results comparable to those of Tables 2 and 3 also for these structures, for which we had not been involved in the data collection and original structure determination.

The dependence of the REGMEAN results on the precision of the input structure bundle was investigated

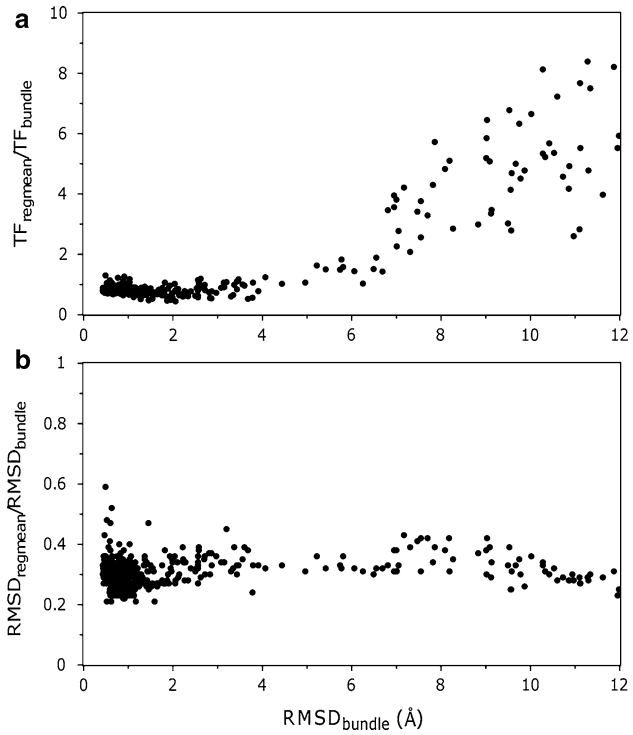

Fig. 5 Dependence of regmean structures for the protein fsh2 on the precision of the corresponding input structure bundles. **a** Target function value of the regmean structure. **b** RMSD of the regmean structure to the mean of the input structure bundle. *Values* on the *vertical axes* are given as ratios relative to the corresponding average values for the input structure bundle. The *horizontal axis* shows the average RMSD of the input conformers to their mean coordinates

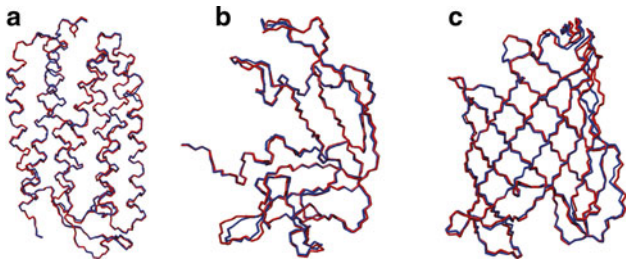

Fig. 6 Application of the REGMEAN algorithm for the transformation of X-ray crystal structures (*blue*) into torsion angle space (*red*). **a** Halorhodopsin (PDB code 1E12) (Kolbe et al. 2000). **b** Bovine pancreatic ribonuclease A (PDB code 3DH5) (Kurpiewska et al. 2009). **c** Protein AT1G79260 from *A. thaliana* (PDB code 3EMM) (Bianchetti et al. 2010). In contrast to the case of NMR structure bundles, experimental restraints were not used in the regularization procedure and positional restraints and torsion angle restraints were referred directly to the single input X-ray structure (instead of the mean coordinates of an input NMR structure bundle)

systematically for the protein fsh2. Input structure bundles of different precision were obtained by randomly omitting part of the distance restraints before the structure calculation. Figure 5 shows that the RMSD of the regmean structures to the mean remains within 20–60% (with few exceptions for high-precision structures even within 20–40%) of the original average values for input structure bundles with average RMSDs to the mean up to 12 Å. However, the target function values stay low only up to RMSDs of the input structure bundle around 5 Å, and increase significantly thereafter. This indicates that the useful "radius of convergence" of the REGMEAN algorithm is about 5 Å. This covers well the range that is relevant for NMR protein structures.

It has been shown that restrained energy refinement in Cartesian space with explicit solvent can improve the quality of NMR structures (Linge et al. 2003). To investigate the influence of such "water refinement" we subjected the regmean structure bundles of Tables 2 and 3 to restrained energy refinement against the AMBER force field (Ponder and Case 2003) using the program OPALp (Koradi et al. 2000; Luginbühl et al. 1996). The results showed that the AMBER energy of the energy-refined regmean structure was always within the range of the corresponding energy-refined input conformers and that, with a relative RMSD of 37–75%, the energy-refined regmean structure remained significantly closer to the mean coordinates of the energy-refined bundle than the average of the individual energy-refined conformers.

The computation time requirements of the REGMEAN algorithm are modest. On a desktop computer with four 3.4 GHz processor cores, the REGMEAN calculations of Tables 2 and 3 took between 12 s for fspo and 148 s for smbp.

Conclusions

With the REGMEAN algorithm we introduce a new method for generating a representative single structure in torsion angle space from an NMR structure bundle, and we propose to modify the way to represent NMR structures of proteins by a 'regmean bundle' that combines the simplicity of a single structure with the ensemble information contained in a structure bundle. While it is standard to represent NMR protein structures by a bundle of conformers rather than by a single structure, X-ray structures are traditionally represented by a single structure. Multimodel handling of crystallographic structural data similar to the handling of NMR ensemble depositions in the PDB has been proposed (Furnham et al. 2006) but has not been widely adopted. Our method bridges the two approaches without discarding the important information on the uncertainty in conformation and data that is contained in an NMR structure bundle and cannot be represented adequately by a single structure. A single structure will always be a lesser interpretation of the information from NMR, as it will not detail the areas of the structure where there is uncertainty.

The REGMEAN algorithm has several features that distinguish it from earlier methods to determine a single representative structure. The regmean structure is calculated in torsion angle space which explicitly excludes distortions of bond lengths, bond angles, and planarities. Instead of trying to reconstruct the proper molecular geometry from the average coordinates, the generation of the regmean structure is as far as possible a standard CYANA structure calculation that starts from scratch and uses the same experimental restraints and the same calculation protocol as the original structure calculation, only supplemented by weak restraints that guide the regmean structure as far as possible towards the center of the original structure bundle. These supplementary restraints are sufficient to obtain a structure in close proximity to the average coordinates, while at the same time being weak enough to allow the experimental restraints to be fulfilled equally well or better than by the conformers of the original structure bundle. Initially, we attempted to achieve a similar result by the simpler approach of computing a large number of interatomic distances from the average coordinates, and applying them, with certain tolerances, as distance restraints in a recalculation of the structure with CYANA. However, the results were clearly inferior to those of the REGMEAN algorithm (data not shown).

The search for an "ideal" single-structure representative of an ensemble of NMR structures has been approached in several ways. In general, the goal of such methods can be summarized as either the identification of an existing structure, in the ensemble in question, which is most representative (Betancourt and Skolnick 2001; Kelley et al. 1996; Zhang and Skolnick 2004), or the generation of a new structure that is structurally close to the average coordinates computed across the ensemble, e.g. by Cartesian space restrained energy minimization of the average structure (Nilges et al. 1988), Newton–Raphson minimization in torsion angle space (Thomas and Pastore 2005), Monte Carlo (Dukka 2009), and procedures from model building (Rotkiewicz and Skolnick 2008). These approaches bear the risk of introducing averaging effects, arising from poorly-defined regions in the structures or from heterogenous ensembles. Several methods have been compared by analyzing the resulting structures with respect to Ramachandran plot and χ^1 distributions (Sutcliffe 1993; Thomas and Pastore 2005), or van-der-Waals clashes (Dukka 2009; Rotkiewicz and Skolnick 2008).

The fact that the regmean structure bundle can be treated as a conventional NMR structure bundle enables its use, without change, in all software packages that can handle structure bundles, and for the deposition of NMR structures in the Protein Data Bank (Berman et al. 2000), a procedure that is not straightforward for more sophisticated approaches to represent NMR structure ensembles, e.g. by reweighted atomic probability densities (Schwieters and Clore 2002).

Finally, it should be noted that the torsion angle space regularization method of the REGMEAN algorithm can readily be used for other applications, including the generation of structural models of point-mutated proteins, the transformation of crystal structures into torsion angle space (Fig. 6), and the application of "soft" rigid body restraints in structure calculations. The latter can be used, for instance, to maintain the conformations of transmembrane helices in NMR structure calculations of membrane proteins with sparse experimental data by treating each transmembrane helix as a "domain" in the sense of the REGMEAN algorithm (Reckel et al. 2011).

Acknowledgments We gratefully acknowledge financial support by the Lichtenberg program of the Volkswagen Foundation and by a Grant-in-Aid for Scientific Research of the Japan Society for the Promotion of Science (JSPS).

Appendix

Bimodal averages and standard deviation of torsion angles

If the values $S = \{\phi_1, \ldots, \phi_n\}$ of a torsion angle ϕ are clustered in two separate regions it makes little sense to determine an average value. Instead it is meaningful to split the set S into two disjoint subsets S_1 and S_2 for the purpose of computing two bimodal average values $\overline{\phi}_1 = \arg\sum_{k\in s_1} e^{i\phi_k}$ and $\overline{\phi}_1 = \arg\sum_{k\in s_2} e^{i\phi_k}$ of the torsion angle values in S_1 and S_2, respectively. The choice of S_1 and S_2 is optimal if it minimizes the bimodal standard deviation

$$\sigma_\phi^{(2)} = \sqrt{\frac{1}{n}\sum_{k=1}^{n} \min\left(|\phi_k - \overline{\phi}_1|, 2\pi - |\phi_k - \overline{\phi}_1|, |\phi_k - \overline{\phi}_2|, 2\pi - |\phi_k - \overline{\phi}_2|\right)^2}$$

that results from summing for each torsion angle value ϕ_k the squared deviation from the closer of the two bimodal average values $\overline{\phi}_1$ and $\overline{\phi}_2$, taking into account the periodicity.

It would be computationally inefficient to evaluate $\sigma_\phi^{(2)}$ for each of the 2^n possible choices of the subsets S_1 and S_2. To determine a good approximation of the optimal bimodal average values in polynomial time, we first calculate the $n \times n$ matrix of torsion angle differences $\Delta\phi_{ij} = \min\left(\left|\phi_i - \phi_j\right|, 2\pi - \left|\phi_i - \phi_j\right|\right)$. For all pairs (i, j) with $\Delta\phi_{ij} > \pi/4$ (to avoid splitting into two hardly separated clusters), we compute $\widetilde{\phi_1} = \arg\sum_{k:\Delta\phi_{ki} \le \Delta\phi_{kj}} e^{i\phi_k}$ and $\widetilde{\phi_2} = \arg\sum_{k:\Delta\phi_{ki} > \Delta\phi_{kj}} e^{i\phi_k}$. (In the exponential functions i denotes the imaginary unit $\sqrt{-1}$, otherwise the index i.) The deviations of the individual torsion angle values ϕ_k from $\widetilde{\phi_1}$ and $\widetilde{\phi_2}$ are given by $\delta_{1k} = \min\left(\left|\phi_k - \widetilde{\phi_1}\right|, 2\pi - \left|\phi_k - \widetilde{\phi_1}\right|\right)$ and $\delta_{2k} = \min\left(\left|\phi_k - \widetilde{\phi_2}\right|, 2\pi - \left|\phi_k - \widetilde{\phi_2}\right|\right)$ for $k = 1, \ldots, n$. The corresponding subsets are $S_1 = \{k|\delta_{1k} \le \delta_{2k}\}$ and $S_2 = \{k|\delta_{1k} > \delta_{2k}\}$. We choose the optimal subsets S_1 and S_2 from the pair (i, j) that yields the largest value of $\left|\sum_{k\in S_1} e^{i\phi_k}\right| + \left|\sum_{k\in S_2} e^{i\phi_k}\right|$ to obtain the bimodal average values $\overline{\phi}_1$ and $\overline{\phi}_2$. If S_2 contains more elements than S_1, we exchange the values of $\overline{\phi}_1$ and $\overline{\phi}_2$ such that $\overline{\phi}_1$ always corresponds to the cluster with the larger number of elements.

References

Antuch W, Güntert P, Wüthrich K (1996) Ancestral $\beta\gamma$-crystallin precursor structure in a yeast killer toxin. Nat Struct Biol 3:662–665

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) Finding the needle in a haystack: educing native folds from ambiguous ab initio protein structure predictions. J Comput Chem 22:339–353

Bianchetti CM, Blouin GC, Bitto E, Olson JS, Phillips GN (2010) The structure and NO binding properties of the nitrophorin-like heme-binding protein from *Arabidopsis thaliana* gene locus At 1g79260.1. Proteins 78:917–931

Bowie JU, Lüthy R, Eisenberg D (1991) A method to identify protein sequences that fold into a known 3-dimensional structure. Science 253:164–170

Calzolai L, Lysek DA, Perez DR, Güntert P, Wüthrich K (2005) Prion protein NMR structures of chickens, turtles, and frogs. Proc Natl Acad Sci USA 102:651–655

Chen VB, Arendall WB, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS, Richardson DC (2010) MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr D 66:12–21

Davis IW, Murray LW, Richardson JS, Richardson DC (2004) MolProbity: structure validation and all-atom contact analysis for nucleic acids and their complexes. Nucleic Acids Res 32:W615–W619

Davis IW, Leaver-Fay A, Chen VB, Block JN, Kapral GJ, Wang X, Murray LW, Arendall WB, Snoeyink J, Richardson JS, Richardson DC (2007) MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. Nucleic Acids Res 35:W375–W383

Dukka BKC (2009) Improving consensus structure by eliminating averaging artifacts. BMC Struct Biol 9:12

Furnham N, Blundell TL, DePristo MA, Terwilliger TC (2006) Is one solution good enough? Nat Struct Mol Biol 13:184–185

Güntert P (2003) Automated NMR protein structure calculation. Prog Nucl Magn Reson Spectrosc 43:105–125

Güntert P (2009) Automated structure determination from NMR spectra. Eur Biophys J 38:129–143

Güntert P, Mumenthaler C, Wüthrich K (1997) Torsion angle dynamics for NMR structure calculation with the new program DYANA. J Mol Biol 273:283–298

Herrmann T, Güntert P, Wüthrich K (2002) Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. J Mol Biol 319:209–227

Hooft RWW, Vriend G, Sander C, Abola EE (1996) Errors in protein structures. Nature 381:272

Horst R, Damberger F, Luginbühl P, Güntert P, Peng G, Nikonova L, Leal WS, Wüthrich K (2001) NMR structure reveals intramolecular regulation mechanism for pheromone binding and release. Proc Natl Acad Sci USA 98:14374–14379

Kainosho M, Torizawa T, Iwashita Y, Terauchi T, Ono AM, Güntert P (2006) Optimal isotope labelling for NMR protein structure determinations. Nature 440:52–57

Kelley LA, Gardner SP, Sutcliffe MJ (1996) An automated approach for clustering an ensemble of NMR-derived protein structures into conformationally related subfamilies. Protein Eng 9:1063–1065

Kirchner DK, Güntert P (2011) Objective identification of residue ranges for the superposition of protein structures. BMC Bioinform 12:170

Kolbe M, Besir H, Essen LO, Oesterhelt D (2000) Structure of the light-driven chloride pump halorhodopsin at 1.8 Å resolution. Science 288:1390–1396

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

Koradi R, Billeter M, Güntert P (2000) Point-centered domain decomposition for parallel molecular dynamics simulation. Comput Phys Commun 124:139–147

Kurpiewska K, Font J, Ribó M, Vilanova M, Lewiński K (2009) X-ray crystallographic studies of RNase A variants engineered at the most destabilizing positions of the main hydrophobic core: further insight into protein stability. Proteins 77:658–669

Laskowski RA, Rullmann JAC, MacArthur MW, Kaptein R, Thornton JM (1996) AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. J Biomol NMR 8:477–486

Linge JP, Williams MA, Spronk CAEM, Bonvin AMJJ, Nilges M (2003) Refinement of protein structures in explicit solvent. Proteins 50:496–506

López-Méndez B, Güntert P (2006) Automated protein structure determination from NMR spectra. J Am Chem Soc 128: 13112–13122

López-Méndez B, Pantoja-Uceda D, Tomizawa T, Koshiba S, Kigawa T, Shirouzu M, Terada T, Inoue M, Yabuki T, Aoki M, Seki E, Matsuda T, Hirota H, Yoshida M, Tanaka A, Osanai T, Seki M, Shinozaki K, Yokoyama S, Güntert P (2004) NMR assignment of the hypothetical ENTH-VHS domain At3g16270 from *Arabidopsis thaliana*. J Biomol NMR 29:205–206

Lüthy R, Bowie JU, Eisenberg D (1992) Assessment of protein models with 3-dimensional profiles. Nature 356:83–85

Luginbühl P, Güntert P, Billeter M, Wüthrich K (1996) The new program OPAL for molecular dynamics simulations and energy refinements of biological macromolecules. J Biomol NMR 8:136–146

Morris AL, Macarthur MW, Hutchinson EG, Thornton JM (1992) Stereochemical quality of protein structure coordinates. Proteins 12:345–364

Nilges M, Clore GM, Gronenborn AM (1988) Determination of three-dimensional structures of proteins from interproton distance data by hybrid distance geometry-dynamical simulated annealing calculations. FEBS Lett 229:317–324

Ohnishi S, Güntert P, Koshiba S, Tomizawa T, Akasaka R, Tochio N, Sato M, Inoue M, Harada T, Watanabe S, Tanaka A, Shirouzu M, Kigawa T, Yokoyama S (2007) Solution structure of an atypical WW domain in a novel β-clam-like dimeric form. FEBS Lett 581:462–468

Pääkkönen K, Tossavainen H, Permi P, Rakkolainen H, Rauvala H, Raulo E, Kilpeläinen I, Güntert P (2006) Solution structures of the first and fourth TSR domains of F-spondin. Proteins 64:665–672

Pantoja-Uceda D, López-Méndez B, Koshiba S, Kigawa T, Shirouzu M, Terada T, Inoue M, Yabuki T, Aoki M, Seki E, Matsuda T, Hirota H, Yoshida M, Tanaka A, Osanai T, Seki M, Shinozaki K, Yokoyama S, Güntert P (2004) NMR assignment of the hypothetical rhodanese domain At4g01050 from *Arabidopsis thaliana*. J Biomol NMR 29:207–208

Pantoja-Uceda D, López-Méndez B, Koshiba S, Inoue M, Kigawa T, Terada T, Shirouzu M, Tanaka A, Seki M, Shinozaki K, Yokoyama S, Güntert P (2005) Solution structure of the rhodanese homology domain At4g01050(175–295) from *Arabidopsis thaliana*. Protein Sci 14:224–230

Pellecchia M, Sem DS, Wüthrich K (2002) NMR in drug discovery. Nat Rev Drug Discov 1:211–219

Ponder JW, Case DA (2003) Force fields for protein simulations. Adv Prot Chem 66:27–85

Reckel S, Gottstein D, Stehle J, Löhr F, Verhoefen MK, Takeda M, Silvers R, Kainosho M, Glaubitz C, Wachtveitl J, Bernhard F, Schwalbe H, Güntert P, Dötsch V (2011) Solution NMR structure of proteorhodopsin. Angew Chem 50:11942–11946

Rosato A, Aramini J, Arrowsmith C, Bagaria A, Baker D, Cavalli A, Doreleijers JF, Eletsky A, Giachetti A, Guerry P, Gutmanas A, Güntert P, F. HY, Herrmann T, Huang YJ, Jaravine V, Jonker HRA, Kennedy MA, Lange OF, Liu G, Malliavin TE, Mani R, Mao B, Montelione GT, Nilges M, Rossi P, van der Schot G, Schwalbe H, Szyperski T, Vendruscolo M, Vernon R, Vranken WF, de Vries S, Vuister GW, Wu B, Yang Y, Bonvin AMJJ (2012) Blind testing of routine, fully automated determination of protein structures from NMR data. Structure 8:227–236

Rosato A, Bagaria A, Baker D, Bardiaux B, Cavalli A, Doreleijers JF, Giachetti A, Guerry P, Güntert P, Herrmann T, Huang YJ, Jonker HRA, Mao B, Malliavin TE, Montelione GT, Nilges M, Raman S, van der Schot G, Vranken WF, Vuister GW, Bonvin AMJJ (2009) CASD-NMR: critical assessment of automated structure determination by NMR. Nat Methods 6:625–626

Rotkiewicz P, Skolnick J (2008) Fast procedure for reconstruction of full-atom protein models from reduced representations. J Comput Chem 29:1460–1465

Schwieters CD, Clore GM (2002) Reweighted atomic densities to represent ensembles of NMR structures. J Biomol NMR 23: 221–225

Scott A, Pantoja-Uceda D, Koshiba S, Inoue M, Kigawa T, Terada T, Shirouzu M, Tanaka A, Sugano S, Yokoyama S, Güntert P (2004) NMR assignment of the SH2 domain from the human feline sarcoma oncogene FES. J Biomol NMR 30:463–464

Scott A, Pantoja-Uceda D, Koshiba S, Inoue M, Kigawa T, Terada T, Shirouzu M, Tanaka A, Sugano S, Yokoyama S, Güntert P (2005) Solution structure of the Src homology 2 domain from the human feline sarcoma oncogene Fes. J Biomol NMR 31: 357–361

Sippl MJ (1993) Recognition of errors in 3-dimensional structures of proteins. Proteins 17:355–362

Sutcliffe MJ (1993) Representing an ensemble of NMR-derived protein structures by a single structure. Protein Sci 2:936–944

Thomas D, Pastore A (2005) WHEATSHEAF: an algorithm to average protein structure ensembles. Acta Crystallogr D 61: 112–116

Wallner B, Elofsson A (2003) Can correct protein models be identified? Protein Sci 12:1073–1086

Wimmer R, Herrmann T, Solioz M, Wüthrich K (1999) NMR structure and metal interactions of the CopZ copper chaperone. J Biol Chem 274:22597–22603

Zhang Y, Skolnick J (2004) SPICKER: a clustering approach to identify near-native protein folds. J Comput Chem 25:865–871

Zhao DQ, Jardetzky O (1994) An assessment of the precision and accuracy of protein structures determined by NMR: dependence on distance errors. J Mol Biol 239:601–607