PREDITOR: a web server for predicting protein torsion angle restraints

Mark V. Berjanskii[1], Stephen Neal[2] and David S. Wishart[1,2,3,*]

[1]Department of Computing Science, [2]Department of Biological Sciences, University of Alberta, Edmonton, AB, Canada and [3]NRC National Institute for Nanotechnology (NINT), Edmonton, AB, Canada

ABSTRACT

Every year between 500 and 1000 peptide and protein structures are determined by NMR and deposited into the Protein Data Bank. However, the process of NMR structure determination continues to be a manually intensive and time-consuming task. One of the most tedious and error-prone aspects of this process involves the determination of torsion angle restraints including phi, psi, omega and chi angles. Most methods require many days of additional experiments, painstaking measurements or complex calculations. Here we wish to describe a web server, called PREDITOR, which greatly accelerates and simplifies this task. PREDITOR accepts sequence and/or chemical shift data as input and generates torsion angle predictions (with predicted errors) for phi, psi, omega and chi-1 angles. PREDITOR combines sequence alignment methods with advanced chemical shift analysis techniques to generate its torsion angle predictions. The method is fast (<40 s per protein) and accurate, with 88% of phi/psi predictions being within 30° of the correct values, 84% of chi-1 predictions being correct and 99.97% of omega angles being correct. PREDITOR is 35 times faster and up to 20% more accurate than any existing method. PREDITOR also provides accurate assessments of the torsion angle errors so that the torsion angle constraints can be readily fed into standard structure refinement programs, such as CNS, XPLOR, AMBER and CYANA. Other unique features to PREDITOR include dihedral angle prediction via PDB structure mapping, automated chemical shift re-referencing (to improve accuracy), prediction of proline cis/trans states and a simple user interface. The PREDITOR website is located at: http://wishart. biology.ualberta.ca/preditor.

INTRODUCTION

The generation of torsion angle restraints is among the most critical steps in protein structure determination by NMR. In many cases, torsion angles play a key role in defining or 'tightening' the secondary structure of protein structures during the structure refinement process. Torsion angles also are used in lieu of other restraints (i.e. NOEs) when these data are missing. The importance of torsion angle information tends to increase with the size of the protein being studied as the quality and quantity of other restraints, such as NOEs, deteriorate due to increased spectral overlap and reduced sensitivity. Because of their importance in structure calculations, all commonly used software packages for NMR structure determination, such as CNS (1), XPLOR (2), CYANA (3) and AMBER (4), accept torsion angles as restraints.

In NMR, the information about torsion angles is commonly obtained from scalar couplings (e.g. $^3J_{HNH\alpha}$, $^3J_{H\alpha\text{-}1N}$, $^3J_{C'H\alpha}$) and cross-correlated relaxation experiments (5–10), and often involves comparing peak intensities or measuring peak splitting. The accuracy of these measurements can be severely compromised by signal broadening and low signal-to-noise ratios, especially when dealing with larger (>150 residue) proteins. Chemical shifts offer an alternative route for obtaining torsion angle restraints. Indeed, it is well known that $^1H\alpha$, $^{13}C\alpha$ and ^{13}CO shifts are very sensitive to backbone ϕ/ψ angles, while ^{15}N shifts appear to be significantly influenced by side chain χ_1 angles of the preceding residue (11). Chemical shift measurements are less affected by peak overlap or reduced spectral sensitivity than measurements of peak intensities and peak splittings. Furthermore, chemical shift measurements are routinely obtained for all peptides and proteins and often do not require additional NMR experiments beyond those needed for backbone assignments. Indeed, chemical shift measurements are commonly done as the first step in the NMR-assisted determination of protein structures.

The simplicity and accuracy of chemical shift measurements in combination with the public availability of thousands of protein chemical shift assignments has prompted the development of a number of protocols that use chemical

---

*To whom correspondence should be addressed. Tel: 780 492 0383; Fax: 780 492 1071; Email: david.wishart@ualberta.ca

shifts to predict backbone dihedral angles (12–17). TALOS (17) is one such example and is currently one of the most commonly used programs that utilize chemical shifts to obtain dihedral angles. TALOS predicts ϕ and ψ torsion angles by comparing the chemical shifts and primary sequence of the protein being studied with a database of homologous polypeptides with known torsion angles and chemical shifts. TALOS has made a profound impact on protein NMR, having appeared in more than 700 published works to date. However, TALOS has certain limitations. For instance, it does not predict side chain (χ_1) or cis-trans peptide bond angles (ω) nor can it handle mis-referenced chemical shifts. Furthermore, TALOS does not provide information about bounds of torsion angle restraints that are required for NMR structure calculations. It also runs very slowly (22 min for a 150 residue protein on a 2.6 GHz processor).

The motivation for the current work was our belief that the prediction of torsion angles could be extended to angles other than ϕ and ψ, that these predictions could be performed much faster, and that their accuracy could be significantly improved if the recent advances in our understanding of chemical shifts and the growing body of protein structural information could be implemented in a prediction protocol. Here we describe a program, called PREDITOR (PREDIction of TORsion angles from chemical shift and homology), that is able to accurately predict a large number of protein torsion angles (ϕ, ψ, ω, χ) using either ^{1}H, ^{13}C and ^{15}N chemical shift assignments or protein sequence (alone) as input. Overall, the program is 35× faster and the accuracy (for combined shift-based and homology-based prediction of ϕ/ψ angles, using a 30° tolerance) is ~20% greater than TALOS. The program can also predict g+, g−, trans states of χ_1 angles with 84% accuracy and ω angles are predicted with essentially 100% accuracy (using a 2-state, cis/trans state prediction). PREDITOR is able to extend the limits of torsion angle prediction accuracy by (i) combining shift based and homology based predictions, (ii) taking advantage of better understanding the relationship between chemical shifts and torsion angles and (iii) utilizing recent advances in correcting mis-referenced NMR assignments and predicting of protein flexibility from chemical shifts. The PREDITOR website is located at: http://wishart.biology.ualberta.ca/preditor.

PROGRAM DESCRIPTION

PREDITOR is composed of two parts, a front-end web interface (Figure 1) written in Python and HTML and a back-end 'calculator' that consists of several programs including RCI (18), CSI (19), VADAR (20), BLAST (21) and REFCOR [based on (22)] as well as several parsing and conversion utilities for reading input files. Four of the programs—BLAST, VADAR, CSI and the core of the PREDITOR code are written in ANSI standard C. Several other programs including RCI, REFCOR and most input parsing and conversion utilities are written in Python. PREDITOR also accesses two databases, a local database of NMR assignments and torsion angles and the Protein Data Bank (23). The source code for the basic algorithm is available from the authors upon request.

PREDITOR accepts three kinds of input files: (i) chemical shift assignments in BMRB NMR-STAR format (24), (ii) chemical shift assignments in SHIFTY format (25) and (iii) raw protein sequence in FASTA format (26). Users can either upload an input file into the web server (via a browse button) or paste the data in a standard text box. Users are also offered several options to adjust program operations to suit their specific needs. By default, PREDITOR uses both chemical shifts (via RefDB) and sequence homology (to structural homologues in the PDB) to predict torsion angles. This usually guarantees maximal performance with minimal user input (27). However, the reliance on homology-based predictions may not be desirable in some cases. For example, a user may wish to assess the agreement between the shift-based (NMR) torsion angles and the torsion angles measured for an existing X-ray structure of the same protein. In such cases, PREDITOR allows users to select the radio-button option to predict dihedral angles from chemical shifts only. In other cases, a user may only want to use homology-based predictions in his or her research. For instance, a comparison of torsion angles predicted from chemical shifts with those predicted via homology may aid in resolving ambiguous cases during the NMR assignment process. In these cases, users may select the radio-button option to predict dihedral angles via homology only. PREDITOR also offer users the flexibility to specify the PDB ID of the protein structure that should be used in homology-based predictions. This feature is especially useful when the structure of the best-matching homologue was solved, say, under significantly different experimental conditions (e.g. high urea concentration to partially unfold the protein) or if the structure corresponds to a protein in a different functional form (e.g. ligand-bound, post-translationally modified, etc.). To help users identify an appropriate homologue, we offer an option to BLAST their protein against the PDB and to display the results so they are hyperlinked to the corresponding pages of the PDB. These options are available only when chemical shifts are submitted to the web server. If a FASTA sequence only is used as input, torsion angles are predicted from a PDB homologue with the best BLAST *E*-value.

One of the more important advantages of PREDITOR over existing programs is its capability to correct mis-referenced chemical shifts (17,27). Chemical shift referencing, particularly for ^{13}C and ^{15}N shifts, continues to be a major problem in biomolecular NMR with about 20% of newly deposited assignments in the BMRB database being mis-referenced (28). Mis-referenced shifts could substantially reduce the performance of chemical shift-based torsion angle predictions. PREDITOR's reference correction option is always turned on by default and can be switched off if necessary.

An average PREDITOR run takes about 38 CPU seconds on a 2.6 GHz processor equipped with 512 Mbytes of RAM. An example of the program output is shown in Figure 1. As might be expected, PREDITOR displays the name of the input file, the options selected, the PDB ID, the BLAST *E*-value and the level of identity of the PDB homologue (if any) used in predictions. Below these data, the predicted torsion angles (ϕ, ψ, ω and χ_1), their errors and confidence scores are shown in a tabular format. These values are also available for download. Torsion angles predicted from a homology model are labeled with asterisk.
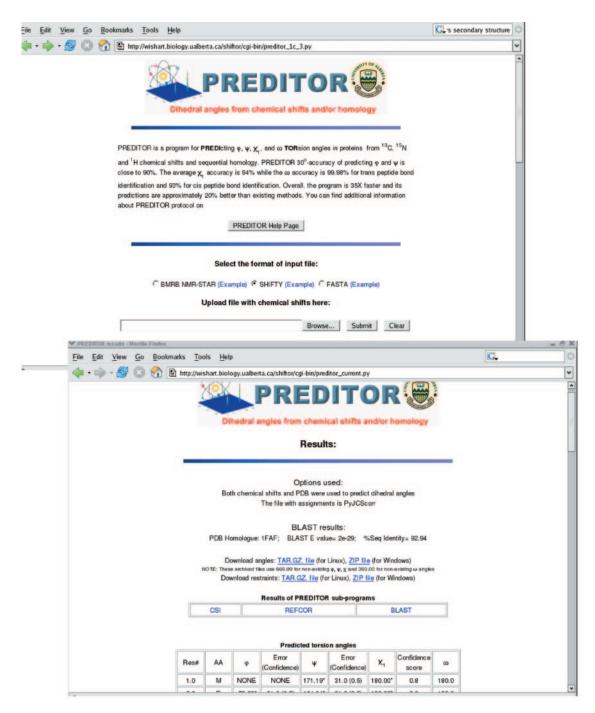
Figure 1. Screenshot montage of PREDITOR's input and output pages. PREDITOR supports BMRB, SHIFTY and FASTA formatted input files

In addition to these data, the output page contains hyperlinks to additional web pages with the results of the sequence alignment ('BLAST results'), chemical shift reference corrections ('REFCOR results'), the predictions of protein flexibility ('RCI results') and secondary structure predictions ('CSI results') derived from chemical shifts. The PREDITOR server also allows users to download the predicted/calculated dihedral angles restraints in CNS, CYANA, AMBER and XPLOR format. In addition to its rich and extensive data output, PREDITOR also offers a comprehensive list of help pages to assist users in preparing their input files, in understanding torsion angle prediction methods and in understanding the program output. This information is provided to make the PREDITOR protocol as transparent as possible and to facilitate any troubleshooting if necessary.

The basic PREDITOR prediction algorithm has been described in much more detail elsewhere (27). What follows is a brief overview of the program's general principles. PREDITOR predicts protein torsion angles using a combined protocol that can be roughly divided into two components: shift-based predictions and homology-based predictions. If a set of NMR assignments is submitted to the program, both

types of predictions are run and their results are merged based on their respective errors or confidence levels, and on 'switch points' as explained below. If a FASTA sequence is supplied as an input file, only homology-based predictions are performed.

Chemical shift-based predictions are initiated by re-referencing the NMR assignments (if this option is not turned off by the user). Shift re-referencing is done by an in-house program called REFCOR. REFCOR applies a recently published protocol that uses Hα chemical shifts for the initial identification of secondary structure and the calculation of reference offsets (22). REFCOR can also go beyond the original procedure and properly predict secondary structure and correct referencing using other nuclei (Cα, CO and Cβ) or a consensus Chemical Shift Index calculation (19). These features enable PREDITOR/REFCOR to re-reference shifts in the absence of Hα assignments. A more detailed description of REFCOR features and performance will be published elsewhere.

For shift-based predictions, PREDITOR derives backbone torsion angles by comparing the chemical shifts of successive amino acid triplets from the query sequence with triplets contained in PREDITOR's own shift/structure database. Currently, the most recent version of this shift/structure database consists of 141 different protein entries obtained from RefDB (providing chemical shift data) and the PDB (providing torsion angle data). The torsion angles for each residue were generated using VADAR. The chemical shifts in the database were re-referenced via SHIFTCOR (28). Updates to the database and updates of PREDITOR's performance relative to its standard test sets will be posted on PREDITOR's Help pages at regular intervals.

To calculate the backbone torsion angles from chemical shifts, PREDITOR calculates a sequence/shift similarity score [$S(i,j)$]. For each query triplet 'i' and each database triplet 'j' the similarity score $S(i,j)$ is calculated using the following equation.

$$
\begin{aligned}
S(i,j) = \Sigma\{ & 0.5 * K_{n1}(\text{SeqSim}) \\
& + K_{n2}(|\delta\Delta C\alpha_{i+n} - \Delta\delta C\alpha_{j+n}|) \\
& + K_{n3}(|\delta\Delta C\beta_{i+n} - \delta\Delta C\beta_{j+n}|) \\
& + K_{n4}(|\delta\Delta CO_{i+n} - \delta\Delta CO_{j+n}|) \\
& + K_{n5}(|\delta\Delta H\alpha_{i+n} - \delta\Delta H\alpha_{j+n}|) \\
& + K_{n6}(|\delta\Delta N_{i+n} - \delta\Delta N_{j+n}|) \\
& + K_{n7}(|\delta\Delta HN_{i+n} - \delta\Delta HN_{j+n}|)\}
\end{aligned}
\qquad \mathbf{1}
$$

where Σ sums over the triplet of $n = -1$ to 1, K_{nm} corresponds to empirically determined weighting coefficients (Table 1 on the PREDITOR help page) for each triplet 'n' of each term 'm', SeqSim represents the sequence similarity between each sequence triplet using the SeqSee weight matrix (29). $\delta\Delta X$ is the secondary chemical shift (30) of nucleus X. The similarity score [$S(i,j)$] between the query triplet and database triplet is calculated for all triplets in the database. The ten triplets with the lowest scores are selected and torsion angles for central triplet residues are extracted. The predicted torsion angles are clustered if the difference of the ϕ or ψ angles is less than 15°. The mean ϕ and ψ angles of the cluster with the lowest overall $S(i,j)$ score are used as the

predicted torsion angles for the central residue of the query triplet.

PREDITOR uses probabilistic χ_1 hypersurfaces calculated by Dunbrack (31) to generate an initial set of predicted χ_1 angles based on conformations of ϕ or ψ angles. Each χ_1 angle is predicted to be in one of three states ($-60°$, $+60°$ and 180°) and assigned a confidence score between 0 and 1 (with 1 being most confident). By default, PREDITOR assigns a trans conformation (180°) to ω angles of all non-proline residues. The presence of cis peptide bonds in proline residues is detected via comparison of the $^{13}C\beta$ and $^{13}C\gamma$ shifts (32). If the absolute difference between the $^{13}C\beta$ and $^{13}C\gamma$ shifts is greater than 9 p.p.m. or if Trp, Tyr, Phe, Gly and Cys are in either $i - 1$ or $i + 1$ positions around proline (33) and $^{13}C\beta$, $^{13}C\gamma$ shift difference is between 8 and 9 p.p.m., PREDITOR assigns a cis-ω angle (0°) to the proline residue.

PREDITOR's homology-based predictions are initiated from a pairwise sequence alignment of the input sequence and all known PDB sequences using BLAST (21). The structure with the lowest E-value is retrieved from the PDB and its dihedral angles are calculated by VADAR (20). Torsion angles are mapped to the query sequence using the sequence alignment provided by BLAST. χ_1 angles are not assigned to Gly and Ala residues. In addition, a ϕ angle of $-65°$ is mapped on to Pro regardless of the values of homology-predicted angles.

Prior to merging shift-based and homology-based predictions, the error limits (for ϕ/ψ angles) and confidence scores (for ϕ, ψ, χ_1) angles are calculated for each method as described below. Predictions with higher confidence scores or lower error limits are given precedence over predictions with lower confidence scores or higher error limits. However, if shift-derived torsion angles of four or more consecutive residues are significantly different (>60°) from PDB-derived values, the shift-based predictions are given precedence. When homology-based predictions are not possible, the shift-derived predictions are used. The selection of either homology-based predictions or shift-based predictions is also determined by 'switch points' implemented in the program. The rationale for adding the switch points was our observation that accuracy of predicted angles decreases significantly with increase of beta-sheet content in proteins and reduction of sequence identity of the homologue identified by BLAST (27). These data are shown in PREDITOR's help pages. The switch points were obtained empirically by thorough analysis of these dependencies and are listed in Web Table 2 on PREDITOR's help page.

PREDITOR assigns an error to each predicted ϕ/ψ torsion angle by combining its confidence scores with predicted or identified secondary structures and local sequence identity. The relationship to confidence scores and estimated ϕ/ψ torsion angle errors are listed in Web Table 3 of PREDITOR's Help page. These error values were determined empirically by attempting to minimize both the size of the assigned error and the number of erroneous predictions with high (≥ 0.7) confidence values. Generally speaking residues in helices have smaller errors than beta strands and ϕ angles have smaller errors than ψ angles. Confidence scores in PREDITOR are calculated using the following formula:

$$
C = 2^{-(3*S(i,j)/239)**2} \qquad \mathbf{2}
$$

Where C is the confidence score and S(i,j) is the similarity score calculated in Equation 1. The confidence scores depend on both the sequence and chemical shift similarity of the query protein sequence triplets to the corresponding sequence triplets found in the PREDITOR database. This formula produces values that range from 0.0 to 1.0 (with 1.0 having the highest confidence). Roughly speaking a PREDITOR confidence rating ≥0.7 corresponds to the TALOS rating of 'good', a rating below 0.4 corresponds to a TALOS rating of 'bad' and between 0.4 and 0.6 is considered 'ambiguous'. With the homology search turned off, PREDITOR predictions having a confidence score ≥0.7, have an error rate that is slightly >3%. Typically 60–65% of PREDITOR predictions (with the homology search turned off) have a confidence score ≥0.7. This is essentially identical to the performance reported for TALOS (17). With PREDITOR's PDB-homology search turned on, the error rate is similar (~3%) but the percentage of accepted predictions is ~15% greater. Likewise, for PDB-based predictions the error limits are generally several degrees smaller. Note also that when PDB homologues are found the confidence scores vary with the local sequence identity of the PDB homologue (see PREDITOR's Help page for more details).

PREDITOR uses a probabilistic hypersurface model (34) to estimate the confidence of χ_1 prediction from chemical shifts. As with the ϕ/ψ torsion angles the confidence values for χ_1 angle predictions vary between 0.0 and 1.0, with 1.0 being the highest level. Confidence levels of homology-based predictions of χ_1 have been derived empirically from the dependence of prediction accuracy on the level homologue sequence identity. Confidence levels derived from PDB homologues are shown in Web Table 4 on PREDITOR's Help page. When assignments for fewer than six nuclei (Cα, Cβ, CO, Hα, N and HN) are available, the predicted torsion angle error is multiplied by a scaling factor (see Web Table 5 on PREDITOR's Help page). These scaling factors were determined for each of the 63 possible assignment combinations by measuring the average increase of PREDITOR's prediction errors for each of these combinations in the PREDITOR database.

VALIDATION

PREDITOR was optimized, tested and evaluated on a training set of 141 proteins (20 489 residues), for which complete or nearly complete ^{1}H, ^{13}C and ^{15}N chemical shifts were known and for which high quality PDB structures were available (Web Table 6 on the PREDITOR help page). To evaluate the performance of the algorithm for the training set, we used a leave-one-out procedure by removing a query protein from PREDITOR database prior to running the algorithm. The accuracy of the predictions, $\Delta(\phi/\psi)$, was determined using the following equation:

$$\Delta(\phi/\psi)_i = (|\phi_{obs} - \phi_{pred}|) + (|\psi_{obs} - \psi_{pred}|) \qquad 3$$

where ϕ_{obs} and ψ_{obs} are observed and ϕ_{pred} and ψ_{pred} are predicted ϕ and ψ angles, respectively. If the $\Delta(\phi/\psi)$ value for residue i was <30° (denoted as ϕ/ψ[30]) the prediction was considered 'correct'. To estimate accuracy χ_1 angles, the observed and predicted χ_1 angles were grouped into trans, gauche+ and gauche− these three categories. Predictions are considered to be correct if both predicted and observed χ_1 angles belonged to the same rotamer group. The overall performance for $\Delta(\phi/\psi)$ or χ_1 was determined by calculating percentage of correctly identified torsion angles from their total number in a given protein. The measurement of ω angle accuracy was based on a similar protocol and the performance of the program to predict and identify the 15 cis peptide bonds in the training set.

The average ϕ/ψ[30] accuracy for the full version of PREDITOR on this 141 protein training set was 90.3%. The accuracy for PREDITOR using shift-derived predictions alone was 67.4%. The average χ_1 accuracy was 83.8% (versus 64.4% for shift-only derived predictions), while the ω accuracy was 99.98% for trans peptide bond identification and 93% for cis peptide bond identification. To assess PREDITOR's relative performance against TALOS a subset of 31 of the 141 proteins were also analyzed by TALOS. The results of this comparison, both in terms of accuracy and computational speed are summarized in Supplementary Table A (See the Supplementary Data and

Table 1. Performance using the ϕ/ψ[30] score of the full version of PREDITOR, a disabled version of PREDITOR (shift-based predictions only) and TALOS for the test set of 15 randomly chosen proteins

BMRB#	Protein name	PDB	% β sheet	ϕ/ψ30 score TALOS	ϕ/ψ30 score preditor (shift only)	ϕ/ψ30 score preditor (full)	TALOS time (s)	Preditor time (s)
6357	Beta-lactamase TEM	1YT4	19.62	0.73	0.74	0.87	2194.00	62
4032	Ribonuclease A	1KF3	43.55	0.63	0.83	0.94	1038.00	28
4364	Stromelysin	1HY7	25.6	0.65	0.78	0.91	1130.00	37
6391	MMP-12, catalytic domain	1Y93	22.15	0.67	0.81	0.90	1372.00	38
4661	APAF-1	1CY5	0	0.82	0.80	0.85	720.90	22
5393	CAP, N-terminal domain	1S0P	1.7	0.78	0.81	0.81	1427.36	39
6136	L-Arabinose binding protein	8ABP	25.57	0.74	0.74	0.98	2285.00	69
5540	Flavodoxin	1F4P	29.25	0.70	0.73	0.76	1218.04	33
4472	BeFx-activated CheY	1JBE	22.66	0.70	0.70	0.99	1019.10	29
4717	FtsZ binding domain of ZipA	1F46	30.71	0.74	0.72	0.85	1161.00	32
5514	GFPuv	1KYS	55.51	0.47	0.53	0.98	1832.00	59
6338	At1g77540	1×MT	34.74	0.67	0.60	0.60	860.05	23
4834	Peptide deformylase	1LM4	34.78	0.65	0.61	0.86	1517.67	43
5891	Hyperthermophile DNA-binding Protein	1R7J	10	0.83	0.89	0.90	771.00	20
4831	Hen egg white lysozyme	3LZT	15.5	0.60	0.71	0.86	1029.42	29
			Average					
			24.76	*0.69*	*0.73*	*0.87*	*1305.04*	*37.53*

http://wishart.biology.ualberta.ca/preditor/help/tablea.html). These data indicate that PREDITOR is substantially faster (37×) and 20.2% more accurate than TALOS.

To test the program further and to ensure that the PREDITOR algorithms had not been over-trained or become biased. an independent test set of 15 randomly chosen proteins (2665 residues) not found in either the TALOS or PREDITOR training set was analyzed by both PREDITOR and TALOS. As seen in Table 1, PREDITOR is 18% more accurate than TALOS for ϕ and ψ predictions, when both PDB-derived and shift-based predictions are used. When only chemical shifts are used PREDITOR is 4% more accurate than TALOS. The average PREDITOR run takes 37.5 CPU seconds while the average TALOS run is 1305 CPU seconds on the same 2.6 GHz processor. These results are essentially identical to the results seen in Supplementary Table A.

In summary, PREDITOR is a web server that is capable to rapidly obtain accurate estimates of ϕ, ψ, χ_1 and ω torsion angles, including their error limits and confidence levels, using only chemical shift assignments or protein sequence as input data. Comparisons suggest that these estimates are as good or better than what can be obtained using existing methods and that the approach used here is generally applicable to any protein for which ^{1}H, ^{13}C and ^{15}N shift assignments are available. In addition to its speed and high level of performance PREDITOR also offers other unique features including dihedral angle prediction via PDB structure mapping, automated chemical shift re-referencing (to improve accuracy), prediction of proline cis/trans states, automated generation of XPLOR, CYANA, AMBER or CNS torsion angle restraint output and a simple-to-use web-based user interface. Because of its improved accuracy and extended capabilities, we believe PREDITOR may lead to important changes in general NMR structure determination protocols, allowing more users to place tighter constraints and/or more weight on torsion angle restraint data.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

Conflict of interest statement. None declared.

REFERENCES

1. Brünger,A.T., Adams,P.D., Clore,G.M., Delano,W.L., Gros,P., Grosse-Kunstleve,R.W., Jiang,J.S., Kuszewski,J., Nilges,M., Pannu,N.S. *et al.* (1998) Crystallography and NMR system (CNS): a new software suite for macromolecular structure determination. *Acta. Cryst.*, **D54**, 905–921.
2. Brunger,A.T. (1992) *X-PLOR Version 3.1: A System for X-ray Crystallography and NMR*. http://xplor.csb.yale.edu/xplor/.
3. Guntert,P. (2004) Automated NMR structure calculation with CYANA. *Meth. Mol. Biol.*, **278**, 353–378.
4. Case,D.A., Cheatham,T.E.,IIIrd, Darden,T., Gohlke,H., Luo,R., Merz,K.M.,Jr, Onufriev,A., Simmerling,C., Wang,B. and Woods,R.J. (2005) The Amber biomolecular simulation programs. *J. Comput. Chem.*, **26**, 1668–1688.
5. Madsen,J.C., Sorensen,O.W., Sorensen,P. and Poulsen,F.M. (1993) Improved pulse sequences for measuring coupling constants in ^{13}C, ^{15}N-labeled proteins. *J. Biomol. NMR*, **3**, 239–244.
6. Hu,J.S. and Bax,A. (1997) Determination of phi and chi(1) angles in proteins from C-13-C-13 three-bond J couplings measured by three-dimensional heteronuclear NMR. How planar is the peptide bond? *J. Am. Chem. Soc.*, **119**, 6360–6368.
7. Pellecchia,M., Fattorusso,R. and Wider,G. (1998) Determination of the dihedral angle psi based on J coupling measurements in N-15/C-13-labeled proteins. *J. Am. Chem. Soc.*, **120**, 6824–6825.
8. West,N.J. and Smith,L.J. (1998) Side-chains in native and random coil protein conformations. Analysis of NMR coupling constants and chi1 torsion angle preferences. *J. Mol. Biol.*, **280**, 867–877.
9. Kloiber,K., Schuler,W. and Konrat,R. (2002) Automated NMR determination of protein backbone dihedral angles from cross-correlated spin relaxation. *J. Biomol. NMR*, **22**, 349–363.
10. Carlomagno,T., Bermel,W. and Griesinger,C. (2003) Measuring the chi1 torsion angle in protein by CH-CH cross-correlated relaxation: a new resolution-optimised experiment. *J. Biomol. NMR*, **27**, 151–157.
11. Neal,S., Nip,A.M., Zhang,H. and Wishart,D.S. (2003) Rapid and accurate calculation of protein 1H, 13C and 15N chemical shifts. *J. Biomol. NMR*, **26**, 215–240.
12. Wishart,D.S., Sykes,B.D. and Richards,F.M. (1992) The chemical shift index: a fast and simple method for the assignment of protein secondary structure through NMR spectroscopy. *Biochemistry*, **31**, 1647–1651.
13. Osapay,K. and Case,D.A. (1994) Analysis of proton chemical shifts in regular secondary structure of proteins. *J. Biomol. NMR*, **4**, 215–230.
14. Wishart,D.S. and Nip,A.M. (1998) Protein chemical shift analysis: a practical guide. *Biochem. Cell. Biol.*, **76**, 153–163.
15. Oldfield,E. (1995) Chemical shifts and three-dimensional protein structures. *J. Biomol. NMR*, **5**, 217–225.
16. Beger,R.D. and Bolton,P.H. (1997) Protein phi and psi dihedral restraints determined from multidimensional hypersurface correlations of backbone chemical shifts and their use in the determination of protein tertiary structures. *J. Biomol. NMR*, **10**, 129–142.
17. Cornilescu,G., Delaglio,F. and Bax,A. (1999) Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J. Biomol. NMR*, **13**, 289–302.
18. Berjanskii,M.V. and Wishart,D.S. (2005) A simple method to predict protein flexibility using secondary chemical shifts. *J. Am. Chem. Soc.*, **127**, 14970–14971.
19. Wishart,D.S. and Sykes,B.D. (1994) Chemical shifts as a tool for structure determination. *Meth. Enzymol.*, **239**, 363–392.
20. Willard,L., Ranjan,A., Zhang,H., Monzavi,H., Boyko,R.F., Sykes,B.D. and Wishart,D.S. (2003) VADAR: a web server for quantitative evaluation of protein structure quality. *Nucleic Acids Res.*, **31**, 3316–3319.
21. Altschul,S.F., Madden,T.L., Schaffer,A.A., Zhang,J., Zhang,Z., Miller,W. and Lipman,D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.
22. Wang,Y. and Wishart,D.S. (2005) A simple method to adjust inconsistently referenced 13C and 15N chemical shift assignments of proteins. *J. Biomol. NMR*, **31**, 143–148.
23. Berman,H.M., Westbrook,J., Feng,Z., Gilliland,G., Bhat,T.N., Weissig,H., Shindyalov,I.N. and Bourne,P.E. (2000) The Protein Data Bank. *Nucleic Acids Res.*, **28**, 235–242.
24. Seavey,B.R., Farr,E.A., Westler,W.M. and Markley,J.L. (1991) A relational database for sequence-specific protein NMR data. *J. Biomol. NMR*, **1**, 217–236.
25. Wishart,D.S., Watson,M.S., Boyko,R.F. and Sykes,B.D. (1997) Automated 1H and 13C chemical shift prediction using the BioMagResBank. *J. Biomol. NMR*, **10**, 329–336.
26. Pearson,W.R. and Lipman,D.J. (1988) Improved tools for biological sequence comparison. *Proc. Natl Acad. Sci. USA*, **85**, 2444–2448.

27. Neal,S., Berjanskii,M., Zhang,H. and Wishart,D.S. (2006) Accurate prediction of protein torsion angles using chemical shifts and sequence homology. *Magn. Reson. Chem.*, (in press).

28. Zhang,H., Neal,S. and Wishart,D.S. (2003) RefDB: a database of uniformly referenced protein chemical shifts. *J. Biomol. NMR*, **25**, 173–195.

29. Wishart,D.S., Boyko,R.F., Willard,L., Richards,F.M. and Sykes,B.D. (1994) SEQSEE: a comprehensive program suite for protein sequence analysis. *Comput. Appl. Biosci.*, **10**, 121–132.

30. Wishart,D.S., Bigam,C.G., Holm,A., Hodges,R.S. and Sykes,B.D. (1995) ^{1}H, ^{13}C and ^{15}N random coil NMR chemical shifts of the common amino acids. I. Investigations of nearest-neighbor effects. *J. Biomol. NMR*, **5**, 67–81.

31. Canutescu,A.A., Shelenkov,A.A. and Dunbrack,R.L.,Jr (2003) A graph-theory algorithm for rapid protein side-chain prediction. *Protein Sci.*, **12**, 2001–2014.

32. Schubert,M., Labudde,D., Oschkinat,H. and Schmieder,P. (2002) A software tool for the prediction of Xaa-Pro peptide bond conformations in proteins based on 13C chemical shift statistics. *J. Biomol. NMR*, **24**, 149–154.

33. Pahlke,D., Freund,C., Leitner,D. and Labudde,D. (2005) Statistically significant dependence of the Xaa-Pro peptide bond conformation on secondary structure and amino acid sequence. *BMC Struct. Biol.*, **5**, 8.

34. Dunbrack,R.L.,Jr (1999) Comparative modeling of CASP3 targets using PSI-BLAST and SCWRL. *Proteins*, Suppl **3**, 81–87.