

## Protein backbone angle restraints from searching a database for chemical shift and sequence homology

Gabriel Cornilescu, Frank Delaglio & Ad Bax\*

Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892-0520, U.S.A.

Received 10 November 1998; Accepted 6 January 1999

**Key words:**  $\phi$  angle,  $\Psi$  angle, backbone angles, chemical shift, homology of chemical shift, protein structure, sequence homology, TALOS

### Abstract

Chemical shifts of backbone atoms in proteins are exquisitely sensitive to local conformation, and homologous proteins show quite similar patterns of secondary chemical shifts. The inverse of this relation is used to search a database for triplets of adjacent residues with secondary chemical shifts and sequence similarity which provide the best match to the query triplet of interest. The database contains  $^{13}\text{C}^\alpha$ ,  $^{13}\text{C}^\beta$ ,  $^{13}\text{C}'$ ,  $^1\text{H}^\alpha$  and  $^{15}\text{N}$  chemical shifts for 20 proteins for which a high resolution X-ray structure is available. The computer program TALOS was developed to search this database for strings of residues with chemical shift and residue type homology. The relative importance of the weighting factors attached to the secondary chemical shifts of the five types of resonances relative to that of sequence similarity was optimized empirically. TALOS yields the 10 triplets which have the closest similarity in secondary chemical shift and amino acid sequence to those of the query sequence. If the central residues in these 10 triplets exhibit similar  $\phi$  and  $\Psi$  backbone angles, their averages can reliably be used as angular restraints for the protein whose structure is being studied. Tests carried out for proteins of known structure indicate that the root-mean-square difference (rmsd) between the output of TALOS and the X-ray derived backbone angles is about 15°. Approximately 3% of the predictions made by TALOS are found to be in error.

### Introduction

The strong dependence of isotropic chemical shifts on protein structure has long been recognized. In particular, the striking correlation between  $^1\text{H}^\alpha$  chemical shift and secondary structure has been studied extensively (Pastore and Saudek, 1990; Williamson, 1990; Wishart et al., 1991; Ösapay and Case, 1994) and the  $^1\text{H}^N$  shift was found to be sensitive to both hydrogen bonding and secondary structure (Pardi et al., 1983; Williamson, 1990; Wishart et al., 1991). The periodicity of the  $\text{H}^N$  shifts observed in many  $\alpha$ -helical structures, in conjunction with the well-established re-

lation between  $\text{H}^N$  chemical shift and hydrogen bond length (Pardi et al., 1983), suggests that they also contain information on helix bending (Kuntz et al., 1991). Similar correlations between the backbone torsion angles  $\phi$  and  $\Psi$  and the  $^1\text{H}^\alpha$  and  $^1\text{H}^\beta$  chemical shifts have been identified, which appear particularly useful for characterization of turns (Ösapay and Case, 1994).

Although most of the earlier reports on the relation between chemical shift and protein structure focus on  $^1\text{H}^\alpha$  and  $^1\text{H}^N$ , with the advent of heteronuclear isotopic enrichment additional chemical shifts have become accessible and offer the potential to make the relation between chemical shift and structure more quantitative. The secondary  $^{13}\text{C}^\alpha$  and  $^{13}\text{C}^\beta$  chemical shifts of a given residue were found to correlate closely with its  $\phi$  and  $\Psi$  torsion angles (Ando et al., 1984; Saito, 1986; Spera and Bax, 1991), and thereby also with secondary structure (Wishart et al., 1991).

\*To whom correspondence should be addressed. E-mail: bax@nih.gov

Supplementary material available from the authors: One figure, showing predicted and X-ray crystal structure torsion angles as a function of residue for human thioredoxin.

Methods have been developed to obtain backbone torsion angle restraints and secondary structure information from either  $^1\text{H}^\alpha$  and  $^{13}\text{C}^\alpha$  (Luginbühl et al., 1995), or  $^{13}\text{C}^\alpha$ ,  $^{13}\text{C}^\beta$ ,  $^{13}\text{C}'$ , and  $^1\text{H}^\alpha$  (Wishart and Sykes, 1994). The empirical correlation between  $\phi$  and  $\Psi$  backbone torsion angles and the  $^{13}\text{C}^\alpha$  and  $^{13}\text{C}^\beta$  chemical shifts also was found useful for identification of N-terminal helix-capping boxes (Gronenborn and Clore, 1994). This same group also introduced an effective method for incorporating the empirical secondary  $^{13}\text{C}^\alpha$  and  $^{13}\text{C}^\beta$  chemical shift profiles into the structure calculation protocol (Kuszewski et al., 1995; Celda et al., 1995). Ab initio calculations (de Dios and Oldfield, 1993) confirm that the backbone  $\phi$  and  $\Psi$  torsion angles strongly affect  $^{13}\text{C}^\alpha$  and  $^{13}\text{C}^\beta$  shielding, and the use of experimental  $^{13}\text{C}^\alpha$ ,  $^{13}\text{C}^\beta$  and  $^1\text{H}^\alpha$  shifts, in conjunction with residue-specific chemical shift surfaces from ab initio methods, has been proposed as a tool for structure refinement (Pearson et al., 1995). Beger and Bolton (1997) proposed an approach to obtain the most probable  $\phi$  and  $\Psi$  angles from correlation maps between backbone chemical shifts of  $^{13}\text{C}^\alpha$ ,  $^{13}\text{C}^\beta$ ,  $^1\text{H}^\alpha$ ,  $^1\text{H}^\text{N}$  and  $^{15}\text{N}$  of a given residue and its backbone torsion angles. They also showed that this information considerably improves structural quality when used in cases where only a very small number of NOE restraints is available.

The similarity in secondary chemical shifts in homologous proteins also has been well recognized (Redfield and Robertson, 1991). Wishart et al. (1997) developed an elegant approach to utilize this similarity during the resonance assignment process. However, a minimum of ca. 30% sequence identity is quoted as the requirement for making this procedure reliable.

Here, we describe a hybrid approach which utilizes both sequence and chemical shift homology to predict the most likely backbone angles for a given residue. The idea is based on the notion that if a string of adjacent amino acids shows high similarity in secondary chemical shifts with a string of amino acids in a database, the central residues in the two strings are likely to have similar backbone torsion angles. In particular, when qualitative similarity in the residue types of the two strings is used as an additional criterion, the approach becomes remarkably robust. In essence, this is a generalization of the idea that helix-capping boxes can be identified best by combined use of their characteristic patterns of chemical shifts and the residue types involved (Gronenborn and Clore, 1994).

## Materials and methods

A database was created which contains nearly complete  $^{13}\text{C}^\alpha$ ,  $^{13}\text{C}^\beta$ ,  $^{13}\text{C}'$ ,  $^1\text{H}^\alpha$  and  $^{15}\text{N}$  chemical shift assignments of 20 proteins (Table 1), together with the backbone torsion angles  $\phi$  and  $\Psi$ , derived from crystal structures solved at a resolution  $\leq 2.2$  Å (nearly 3000 residues, 14 000 chemical shifts). The format is such that the database can easily be extended by adding new structures for which at least four of the five chemical shifts are available per residue, and for which the structure is known accurately. The structural data follows the Brookhaven Protein Databank (PDB) format and the chemical shifts are in the BioMagResBank (Seavey et al., 1991) format. Residues with missing crystallographic coordinates (e.g. residues 1–17 of cutinase and the amino- and carboxy-terminal residues) as well as residues with multiple conformations in the X-ray structure have been excluded. Residues with high temperature ( $B$ ) factors for the backbone atoms, exceeding 1.5 times the average  $B$ -factor for that protein, were also excluded. This includes the vast majority of cases where differences between crystal and solution structures previously have been noted.

When using collections of chemical shifts of proteins reported by different groups, it is critical to ensure that the same chemical shift referencing convention is used for all these proteins. This is particularly important for  $^{13}\text{C}$  and  $^{15}\text{N}$ , where a wide variety of direct and indirect referencing methods have been used. Rather than relying on the information supplied with the deposited chemical shift data, we evaluate the need for applying a correction to  $^{13}\text{C}$  shifts by calculating how much, on average, the secondary shifts (calculated by subtracting the random coil shifts of Spera and Bax, 1991) deviate from the corresponding secondary chemical shifts predicted by the  $(\phi,\Psi)$ -surfaces of Spera and Bax. These averages are conveniently calculated with a routine added to the X-PLOR program (Brünger, 1993) by Kuszewski et al. (1995), and intended for use of the secondary  $^{13}\text{C}^\alpha$  and  $^{13}\text{C}^\beta$  shifts during structure calculation. We apply a chemical shift correction only if the average deviation for a given protein exceeds by more than a factor of three the expected random variation in this average [i.e., the standard error of ca. 1 ppm (Spera and Bax, 1991) divided by the square root of the number of shifts used]. This manner of correcting the deposited chemical shifts ensures that all secondary shifts are defined in the same manner, and corresponds

Table 1. Proteins contained in the database

Protein (Chemical shifts ref.) (*BioMagResBank no.)	No. of residues	X-ray structure ref. (*PDB code)	Resolution (Å)	Shifts
Alpha-lytic protease (Davis et al., 1997)	198	Fujinaga et al., 1985 (*2alp)	1.7	C <sup>α</sup> , C <sup>β</sup> , C', H <sup>α</sup> , N
Basic pancreatic trypsin inhibitor (Hansen, 1991)	58	Wlodawer et al., 1984 (*5pti)	1.1	C <sup>α</sup> , C <sup>β</sup> , C', H <sup>α</sup> , N
Calbindin (Drakenberg et al., 1989) (*390)	76	Svensson et al., 1992 (*4icb)	1.6	C <sup>α</sup> , C <sup>β</sup> , H <sup>α</sup> , N
Calmodulin (Ikura et al., 1990) (*547)	148	Chattopadhyaya et al., 1992 (*1cll)	1.7	C <sup>α</sup> , C <sup>β</sup> , C', H <sup>α</sup> , N
Calmodulin/M13 (Ikura et al., 1991) (*1634)	147	Medor et al., 1992 (*1cdl)	2.2	C <sup>α</sup> , C <sup>β</sup> , C', H <sup>α</sup> , N
Cutinase (Pompers et al., 1997) (*4101)	214	Longhi et al., 1997 (*1cez)	1.0	C <sup>α</sup> , C <sup>β</sup> , C', H <sup>α</sup> , N
Cyclophilin (Ottiger et al., 1997)	165	Ke et al., 1991 (*2cpl)	1.63	C <sup>α</sup> , C <sup>β</sup> , H <sup>α</sup> , N
Cyanovirin-N (Bewley et al., 1998)	101	Yang et al., 1999 (*3ezm)	1.5	C <sup>α</sup> , C <sup>β</sup> , C', H <sup>α</sup> , N
Dehydrase (Copie et al., 1996)	171	Leesong et al., 1996 (*1mka)	2.0	C <sup>α</sup> , C <sup>β</sup> , C', H <sup>α</sup> , N
D-maltodextrin-binding protein (Gardner et al., 1998)	370	Sharff et al., 1993 (*1dmb)	1.8	C <sup>α</sup> , C <sup>β</sup> , C', H <sup>α</sup> , N
HIV-1 protease (Yamazaki et al., 1996)	99	Lam et al., 1994	1.8	C <sup>α</sup> , C <sup>β</sup> , C', H <sup>α</sup> , N
Human carbonic anhydrase I (Sethson et al., 1996) (*4022)	260	Kumar and Kannan, 1994 (*1hcb)	1.6	C <sup>α</sup> , C <sup>β</sup> , C', H <sup>α</sup> , N
Human thioredoxin in reduced form (Qin et al., 1996)	105	Weichsel et al., 1996 (*1ert)	1.7	C <sup>α</sup> , C <sup>β</sup> , H <sup>α</sup> , N
III-glc (Pelton et al., 1991)	168	Worthylake et al., 1991 (*1f3g)	2.1	C <sup>α</sup> , C <sup>β</sup> , C', H <sup>α</sup> , N
Interleukin-1 $\beta$ (Clore et al., 1990) (*1061)	153	Veerapandian et al., 1992 (*4i1b)	2.0	C <sup>α</sup> , C <sup>β</sup> , H <sup>α</sup> , N
Metallo- $\beta$ -lactamase (Scrofani et al., 1998) (*4102)	232	Concha et al., 1996 (*1znb)	1.85	C <sup>α</sup> , C <sup>β</sup> , C', H <sup>α</sup> , N
Profilin (Archer et al., 1994)	125	Fedorov et al., 1994 (*1acf)	2.0	C <sup>α</sup> , C <sup>β</sup> , C', H <sup>α</sup> , N
Serine protease PB 92 (Fogh et al., 1995)	269	Betzel et al., 1992 (*1svn)	1.4	C <sup>α</sup> , C <sup>β</sup> , C', H <sup>α</sup> , N
Staphylococcal nuclease (D.A. Torchia, personal communication)	141	Loll and Lattman, 1989 (*1snc)	1.65	C <sup>α</sup> , C <sup>β</sup> , C', H <sup>α</sup> , N
Ubiquitin (Wang et al., 1995)	76	Vijay-Kumar et al., 1987 (*1ubq)	1.8	C <sup>α</sup> , C <sup>β</sup> , C', H <sup>α</sup> , N

The table lists references describing the chemical shifts, the X-ray structure, the accession codes for data deposited in the BMRB and PDB databases, the resolution at which the crystal structure was solved, and the types of nuclei for which chemical shifts are available.

to subtraction of the random coil  $^{13}\text{C}^{\alpha}$  and  $^{13}\text{C}^{\beta}$  shifts of Spera and Bax (1991) and  $^{13}\text{C}'$  (Wishart et al., 1995a) from experimentally determined shifts relative to internal trimethylsilyl propionate (TSP). Note that TSP resonates upfield from the IUPAC-recommended standard (Markley et al., 1998), dimethylsilapentane-5-sulfonic acid or DSS, by an insignificant amount (0.12 ppm at pH 7) (Wishart et al., 1995b). The same correction procedure must be used for all other new proteins added to the database. Only a small fraction of the proteins required the above correction procedure. For  $^{15}\text{N}$ , the chemical shift reference standard is liquid ammonia at 25 °C, and the need for application of a correction was evaluated by calculating the average  $^{15}\text{N}$  chemical shifts for all non-Gly, non-Ser, non-Thr residues in  $\alpha$ -helical and  $\beta$ -strand regions of the protein and comparing them with the database averages (119.47 ppm for  $\alpha$ -helices and 122.38 ppm for  $\beta$ -strands). Whenever the average of the  $\alpha$ -helix and  $\beta$ -strand  $^{15}\text{N}$  chemical shift deviations (weighted according to the number of residues used for each type of

secondary structure) is larger than 1 ppm, a correction to the chemical shifts needs to be applied. Alphalytic protease was the only protein for which such a  $^{15}\text{N}$  chemical shift adjustment (by -2.26 ppm) needed to be used. For  $^1\text{H}$ , where historically chemical shift referencing has been much less of a problem, no such corrections were applied.

To investigate whether the  $^{13}\text{C}'$  chemical shift is strongly influenced by the hydrogen bond length, hydrogens were added to the 1.1 Å crystal structure of basic pancreatic trypsin inhibitor (Wlodawer et al., 1984) with the program X-PLOR (Brünger, 1993). For the 24 carbonyls involved in stable backbone-backbone hydrogen bonds, no significant correlation was found between the lengths of the backbone-backbone hydrogen bonds, calculated from this structure, and the corresponding  $^{13}\text{C}'$  secondary shifts. This result suggests that the  $^{13}\text{C}'$  secondary shift is primarily a function of the backbone geometry, in agreement with its previously reported correlation with secondary structure (Kricheldorf and Muller, 1983; Wishart et al.,

Table 2. Empirically optimized  $k$  factors,  $k_n^m$  ( $m$ : homology,  $C^\alpha$ ,  $N$ ,  $C^\beta$ ,  $C'$ ,  $H^\alpha$ ;  $n = -1, 0, 1$ ), for weighting the relative importance of a given chemical shift or residue type in determining the similarity score,  $S(i, j)$  of Equation 1

Res.	Homology	$^{15}\text{N}$	$^1\text{H}^\alpha$	$^{13}\text{C}'$	$^{13}\text{C}^\alpha$	$^{13}\text{C}^\beta$
$n = -1$	0.74	0.16	14.66	1.15	0.72	0.76
$n = 0$	1.48	0.18	17.54	1.21	0.99	0.91
$n = 1$	0.74	0.20	15.25	1.04	0.72	0.70

1991). Therefore, we decided to include the  $^{13}\text{C}'$  shift information in the evaluation, even while for several proteins no  $^{13}\text{C}'$  shifts have been reported in the database.

Although the  $^{15}\text{N}$  chemical shift is known to be influenced by hydrogen bonding (de Dios et al., 1993), it is also influenced by backbone geometry and therefore is included as an input parameter in the torsion angle prediction procedure. However, as discussed below, optimization of the torsion angle prediction program results in a relatively low weighting factor for this chemical shift.

## Results and discussion

### Description of the search procedure

The backbone torsion angle prediction package TALOS (Torsion Angle Likelihood Obtained from Shifts and sequence similarity) is written in the Tcl/Tk language (Osterhout, 1994) and uses NMRWish, a companion package to the NMRPipe processing and analysis system (Delaglio et al., 1995). NMRWish is a version of the Tcl/Tk script interpreter ‘wish’ (Osterhout, 1994), which has been customized to include a relational database engine for manipulation of spectral information and molecular coordinates. An outline of the prediction method used by TALOS is presented in Figure 1.

TALOS reads the experimental protein chemical shift tables and converts them to secondary chemical shifts before entering them in the database. In its current implementation, TALOS evaluates the similarity in amino acid sequence and secondary shifts for a string of three sequential amino acids relative to all triplets of sequential residues contained in the database. Although we expect that further improvement in performance might be attainable for string lengths longer than three, the number of residues in

the database is presently too small to yield a sufficient sampling for such longer strings. For each query triplet of consecutive residues, the similarity to a triplet with center-residue  $j$  in the database is evaluated by computing a similarity factor,  $S(i, j)$ , given by:

$$S(i, j) = \sum_{n=-1}^{+1} \left[ k_n^0 \Delta_{\text{ResType}}^2 + k_n^1 (\Delta \delta C_{i+n}^\alpha - \Delta \delta C_{j+n}^\alpha)^2 + k_n^2 (\Delta \delta N_{i+n} - \Delta \delta N_{j+n})^2 + k_n^3 (\Delta \delta C_{i+n}^\beta - \Delta \delta C_{j+n}^\beta)^2 + k_n^4 (\Delta \delta C'_{i+n} - \Delta \delta C'_{j+n})^2 + k_n^5 (\Delta \delta H_{i+n}^\alpha - \Delta \delta H_{j+n}^\alpha)^2 \right] \quad (1)$$

and the value of  $S(i, j)$  is evaluated for all triplets  $j$  in the database.  $\Delta \delta$  denotes the secondary shifts of the  $^{13}\text{C}^\alpha$ ,  $^{13}\text{C}^\beta$ ,  $^{13}\text{C}'$ ,  $^1\text{H}^\alpha$  and  $^{15}\text{N}$  nuclei. For Gly residues,  $^1\text{H}^\alpha$  shifts are calculated as the average of  $^1\text{H}^{\alpha 2}$  and  $^1\text{H}^{\alpha 3}$ . Values for the weighting factors  $k_n^0$  through  $k_n^5$  are optimized as described below and are given in Table 2; the residue-type similarity matrix ascribes a number to how similar two types of amino acids are and this  $20 \times 20$  matrix is shown in Table 3. The composition of this similarity matrix is largely based on empirical knowledge that, for example, Gly frequently has a positive  $\phi$  angle, Pro has a very restricted range of  $\phi$  angles, and  $C^\beta$ -branched residues are frequently found in  $\beta$ -sheets. There has been some empirical adjustment of the similarity matrix during the process of optimizing the performance of the TALOS program, but results were not found to be particularly sensitive to small changes (by  $\pm 1$ ) in the Table 3 matrix elements. Using the empirical  $k$  values of Table 2, and  $\Delta_{\text{ResType}}$  of Table 3,  $S(i, j)$  values typically range from 5 to 600.

For all database triplets,  $j$ , that yield a  $S(i, j)$  value lower than an adjustable threshold (typically  $\sim 150$ ), TALOS reports the corresponding X-ray crystal structure  $\phi$  and  $\Psi$  angles of residue  $j$ , together with the  $S(i, j)$  value. The threshold is set sufficiently large to obtain a minimum of at least 10 matches for each residue  $i$ .

Optimization of the 15 chemical shift weighting factors made use of a scheme which finds all triplets of residues in the database for which the central residue has  $\phi/\Psi$  angles within  $15^\circ$  of those of a query residue. We then calculate the average and the standard deviation of the secondary chemical shifts for each of

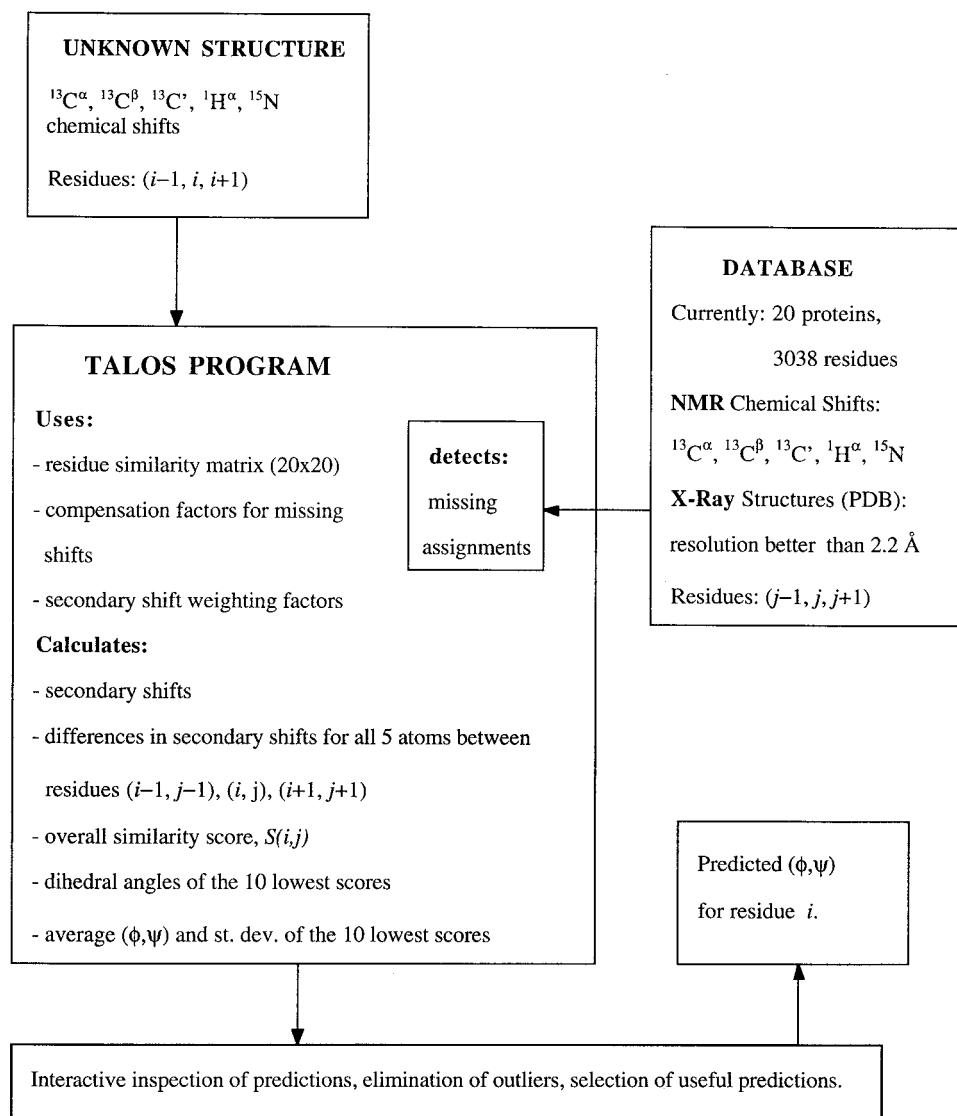


Figure 1. Flow chart of the TALOS program.

the 15 types of chemical shifts (5 nuclei for residue  $i - 1$ ,  $i$ , and  $i + 1$  over this ensemble of triplets. The root-mean-square value of all database secondary chemical shifts of a given type of nucleus, divided by the standard deviation derived in the above described manner, provides a measure for how useful a given type of secondary chemical shift (e.g.,  $\Delta\delta_{i-1}$ ) is at providing information on the  $\phi/\Psi$  angles of residue  $i$ . This ratio was calculated 183 times, each time using a different cutinase residue as the query residue. The chemical shift weighting factors listed in Table 2 are derived from the averages of these respective ratios, after scaling to compensate for the intrinsically differ-

ent widths of the secondary shift distributions of the types of atoms involved (i.e., using the rms values of the  $^{15}\text{N}, ^1\text{H}^\alpha, ^{13}\text{C}^\alpha, ^{13}\text{C}^B$  and  $^{13}\text{C}'$  secondary chemical shift values in the entire database).

The relative weight of the residue type homology versus secondary shifts in the  $S(i, j)$  formula ( $k_n^0$  factors in Equation 1) was optimized empirically, by searching for  $k_n^0$  factors that minimize the number of erroneous predictions, using all residues present in the database for test purposes.

If a particular chemical shift is missing, the corresponding secondary chemical shift difference between the query and the corresponding database chemical

Table 3. Residue similarity factors,  $\Delta_{\text{ResType}}$ , used by TALOS in Equation 1

Res.	A	R	N	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	V
	D																	
A	0	1	1	1	1	1	2	1	2	1	1	1	2	3	1	2	2	2
R	1	0	1	1	1	1	2	1	2	1	0	1	1	3	1	2	1	2
N/D	1	1	0	1	1	1	2	1	2	1	1	1	1	3	1	2	1	2
C	1	1	1	0	1	1	2	1	2	1	1	1	1	3	1	2	1	2
Q	1	1	1	1	0	1	2	1	2	1	1	1	1	3	1	2	1	2
E	1	1	1	1	1	0	2	1	2	2	2	1	1	3	1	2	1	2
G	2	2	2	2	2	2	0	3	3	3	3	3	3	3	3	3	3	3
H	1	1	1	1	1	1	3	0	2	1	2	2	1	3	2	2	1	2
I	2	2	2	2	2	2	3	2	0	1	2	2	2	3	2	1	2	0
L	1	1	1	1	1	2	3	1	1	0	1	1	1	3	2	2	1	2
K	1	0	1	1	1	2	3	2	2	1	0	1	2	3	1	2	2	2
M	1	1	1	1	1	1	3	2	2	1	1	0	2	3	1	2	2	2
F	2	1	1	1	1	1	3	1	2	1	2	2	0	3	2	2	0	1
P	3	3	3	3	3	3	3	3	3	3	3	3	3	0	3	3	3	3
S	1	1	1	1	1	1	3	2	2	2	1	1	2	3	0	1	2	2
T	2	2	2	2	2	2	3	2	1	2	2	2	2	3	1	0	1	1
W/Y	2	1	1	1	1	1	3	1	2	1	2	2	0	3	2	1	0	1
V	2	2	2	2	2	2	3	2	0	2	2	2	1	3	2	1	1	0

shift is set to 1.5 times the rms value of the corresponding secondary chemical shift (rms values are 4.56 ppm for  $^{15}\text{N}$ , 2.49 ppm for  $^{13}\text{C}^\alpha$ , 0.51 ppm for  $^1\text{H}^\alpha$ , 2.01 ppm for  $^{13}\text{C}^\beta$ , and 2.02 ppm for  $^{13}\text{C}'$ ). This way of dealing with incomplete assignments decreases the likelihood that database residues with incomplete assignments contribute to the  $(\phi, \Psi)$  output of TALOS, but does not exclude them altogether.

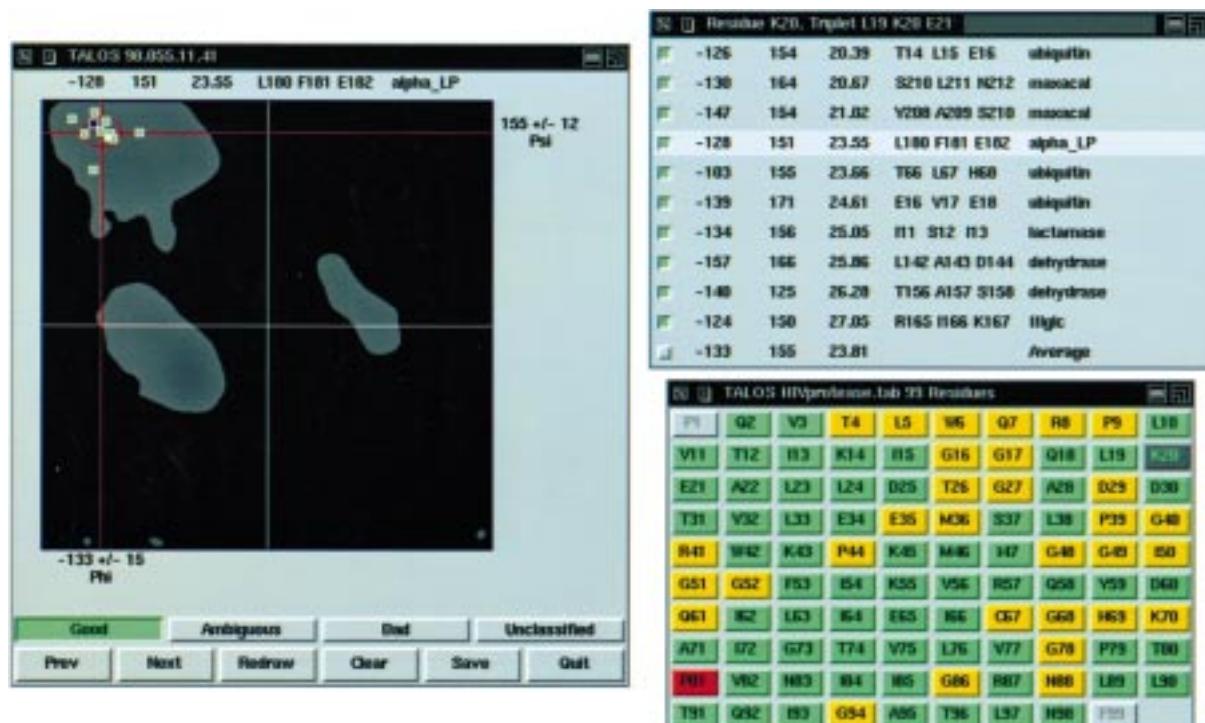
#### Torsion angle prediction

To date, the database used by TALOS contains only 20 structures for which both a high-resolution X-ray structure and nearly complete resonance assignments are available. The reason we felt it is not warranted to include proteins for which a high-resolution NMR structure but no crystal structure is available is that, as discussed below, the agreement between the  $\phi$  and  $\Psi$  angles of most NMR structures and the output of TALOS is considerably lower than for the high-resolution crystal structures in the database.

The TALOS output for the  $\phi$  and  $\Psi$  backbone angles of the center residue in each string consists of the average of the corresponding angles in the 10 strings in the database with the highest degree of similarity (cf. Equation 1). In a first, fully automated but very conservative mode of analysis, the program classifies only those predictions for which at least 9 out of 10 predic-

tions fall in the same populated (gray shaded) region of the Ramachandran map (Figure 2), and none of the center residues in the 10 strings has a positive  $\phi$  angle. If a single residue falls well outside the Ramachandran region in which the remaining 9 residues are located, its  $\phi/\Psi$  values are excluded from calculating the average and rmsd. This procedure typically results in predictions for only about 40% of the residues.

A subsequent interactive inspection of the results, using the graphical interface described below, permits additional predictions to be made. For example, if several predictions fall just outside the most populated region of the Ramachandran map, but generally cluster well with the other  $\phi/\Psi$  predictions, the prediction should be accepted. In some cases, there is one center residue in the ensemble of 10 most similar triplets for which either  $\phi$  or  $\Psi$  deviates by more than 2 standard deviations from the average value for that angle. Empirical testing indicates that it is safe to remove (at most) one such triplet from the ensemble of 10 (TALOS then recalculates the new average  $\phi$  and  $\Psi$  angles and their rmsd), provided that the outlier does not have its  $\phi$  angle in the  $0^\circ < \phi < +150^\circ$  range, and the average  $S(i, j)$  value is less than 80. When the TALOS output for a given query residue yields a



**Figure 2.** Graphical display of TALOS output for HIV protease. The lower right window shows the amino acid sequence, with predictions for each residue designated as ‘good’ (green), ‘ambiguous’ (yellow), or ‘bad’ (red). The prediction data for the selected residue, K20, are listed in the prediction display (top right) and graphed in the Ramachandran display (left). The 10 individual matches from the database are indicated as small green squares in the Ramachandran display, and for reference purposes, the known  $\phi/\Psi$  position from the HIV protease X-ray structure (blue square) is also shown. Clicking on any of the squares highlights the corresponding triplet in the prediction display.

cluster where at least 9 residues have positive  $\phi$  angles, this prediction also should be accepted.

The standard deviations and the range of  $(\phi, \Psi)$  values in the 10 (or 9) most similar database strings provide a measure for the uncertainty in these averages. When this standard deviation exceeds  $45^\circ$ , the prediction must be deemed ‘ambiguous’, and it is recommended that the result of the prediction not be used without careful further inspection of other data, such as the  $d_{\alpha N}(i-1, i)/d_{\alpha N}(i, i)$  NOE intensity ratio (which provides information on the  $\Psi$  angle), the  $^3J_{HN\alpha}$  coupling ( $\phi$  angle), and  $^1J_{C\alpha H\alpha}$  (primarily for identifying positive  $\phi$  angles; Vuister et al., 1992, 1993). Not including such cases where NOE or J coupling information is needed, the above described protocol typically allows a definitive prediction of the  $\phi$  and  $\Psi$  angles to be made for about two thirds of the residues.

Because the number of proteins for which complete NMR assignments and high resolution crystal structures are available is still very limited, the TALOS database usually contains insufficient entries for

unambiguous identification of residues with positive  $\phi$  angles. However, testing indicates that if the center residue of a query triplet has a positive  $\phi$  angle, this frequently results in a significant fraction of center residues which also have positive  $\phi$  angles in the 10 most similar database triplets. These positive  $\phi$  angle triplets typically yield the lowest  $S(i, j)$  values, suggesting that the program will successfully predict most of the positive  $\phi$  angles once the database becomes sufficiently large. For now, unambiguous identification of such positive  $\phi$  angles in most cases requires additional experimental data, such as a very small  $^1J_{C\alpha H\alpha}$  ( $<136$  Hz) (Vuister et al., 1992, 1993), or the presence of an exceptionally strong intraresidue  $H^N-H^\alpha$  NOE.

#### Display of TALOS output

A graphical interface for inspecting and interactively updating the TALOS output is available. An example of its use is shown in Figure 2 for the HIV protease. The interface consists of three windows: the sequence

display, the prediction display, and the Ramachandran display.

The sequence display lists the residues in the protein whose backbone angles are being predicted. The residues are color-coded according to whether the overall prediction for a given residue was designated as good, ambiguous, or bad. In the initial display, before interactive analysis, residues are color-coded as green (prediction accepted in automated mode) and gray (requires inspection). If the true  $\phi/\Psi$  angles are known, residues for which a wrong prediction was accepted can be classified as bad (red), which is convenient for testing purposes. All residues for which TALOS has made predictions which meet the criteria listed above, are highlighted in green. Residues shaded in yellow are those for which no firm prediction can be made, but which nevertheless may contain useful information. For example, if for a given residue 5 out of the 10 triplets show a positive  $\phi$  angle, this suggests that there is a high likelihood that the center residue of the query triplet has a positive  $\phi$  angle.

When a given residue is selected in the sequence display (K20 in Figure 2), the  $\phi$ ,  $\Psi$ , and  $S(i, j)$  parameters are listed in the prediction display, together with the residue numbers and the names of the proteins from which the triplets were taken. The 10  $\phi/\Psi$  pairs are graphed in the Ramachandran display, which also shows the most populated areas of the entire database, shaded in gray. If a reference or trial structure for the query protein is available, its  $\phi/\Psi$  angles will also be graphed on the Ramachandran display (blue square). By clicking on an individual match in the Ramachandran display, it is possible to include or remove this entry from the overall prediction, which is based on the average and standard deviations of the selected matches.

The final results are summarized in an ASCII text table which gives the average  $\phi/\Psi$  angles and their standard deviations for each residue. Versions of the TALOS program are available for most types of UNIX platforms.

#### *Accuracy of TALOS-predicted angles*

Figure 3 plots the predicted  $\phi$  and  $\Psi$  angles of ubiquitin versus those of the high resolution crystal structure. As can be seen from this plot, TALOS does considerably more than classifying residues by their type of secondary structure, and there is a good correlation between predicted and crystallographic torsion angles, even when considering only the residues with a positive  $\Psi$  angle, for example.

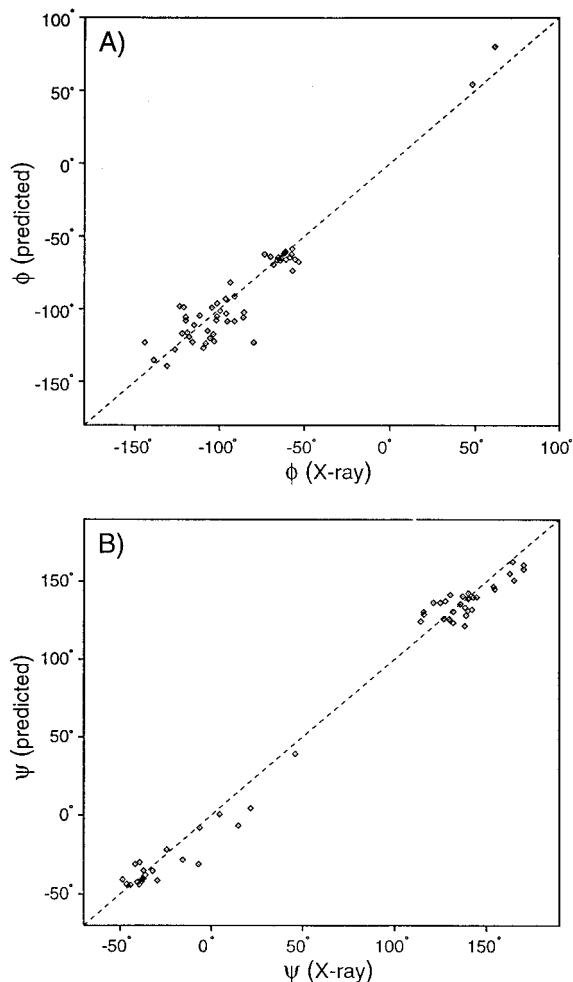


Figure 3. Plots of the backbone angles (A)  $\phi$ , and (B)  $\Psi$  predicted by TALOS, versus those observed in the crystal structure, for ubiquitin.

Figure 4 shows the predicted  $\phi$  and  $\Psi$  angles as a function of residue number, together with the corresponding crystallographically determined angles. The error bars correspond to the standard deviation from the average angle for the center residue of the 10 (or 9) best fitting triplets in the database. No result is shown if this standard deviation exceeds  $45^\circ$ , or if any (but less than 9) of the  $\phi$  angles of the center residues have a positive  $\phi$  angle.

Tests of the accuracy of TALOS predictions were made by eliminating each protein from the database and using the program to predict its backbone angles (Table 4). We found that for about 2% of the residues in the database (i.e., 3% of the predictions made) TALOS predicts the wrong torsion angles. Some examples are:

Table 4. Summary of TALOS results when applied to predicting backbone angles of proteins included in the database

Name	Good (%)	Bad (%)	Ambig. (%)	Avail.	All			
HIV-1 protease	65	67.0	1	1.0	31	32.0	97	99
III-glc	87	61.7	4	2.8	50	35.5	141	168
Alpha-lytic protease	101	54.6	3	1.6	81	43.8	185	198
BPTI	32	58.2	4	7.3	19	34.5	55	58
Calbindin	48	72.7	0	0.0	18	27.3	66	75
Calmodulin	107	84.9	0	0.0	19	15.1	126	148
Calmodulin/M13	98	80.3	0	0.0	24	19.7	122	148
Cutinase	122	66.7	5	2.7	56	30.6	183	214
Cyanovirin-N	55	61.1	1	1.1	34	37.8	90	101
Cyclophilin	87	54.0	5	3.1	69	42.9	161	165
Dehydrase	91	62.7	3	2.1	51	35.2	145	171
HCA I	149	60.3	7	2.8	91	36.9	247	260
Interleukin-1 $\beta$	75	61.0	3	2.4	45	36.6	123	153
Lactamase	137	66.2	4	1.9	66	31.9	207	232
Serine protease PB 92	161	62.7	8	3.1	88	34.2	257	269
D-MBP	217	62.0	5	1.4	128	36.6	350	370
Profilin	72	67.9	0	0.0	34	32.1	106	125
Staphylococcal nuclease	81	67.5	4	3.3	35	29.2	120	141
Human thioredoxin	79	76.7	1	1.0	23	22.3	103	105
Ubiquitin	53	75.7	0	0.0	17	24.3	70	76
Total:								
predictions:	2910							
correct:	1920 (65.3%)							
incorrect:	58 (2.0%)							

Listed are the number of 'Good' predictions, and the percentage relative to the total number of residues with acceptable *B* factors (Avail.), the number of 'Bad' predictions, and the number of residues for which no predictions could be made (Ambig.), plus the total number of residues (All).

(1) Thr<sup>45</sup> in cutinase: predicted  $\Psi = -4 \pm 10^\circ$ ; X-ray  $\Psi = 163^\circ$ . Although the *B* factor is not unusually high, <sup>15</sup>N relaxation data indicate that this residue is located in the middle of a flexible loop which differs in conformation relative to the crystal structure (Prompers et al., 1997).

(2) Asp<sup>159</sup> of beta-hydroxydecanoyl thiol ester dehydrase: predicted  $\phi = -57 \pm 7^\circ$ ,  $\Psi = -36 \pm 10^\circ$ ; X-ray  $\phi = 56^\circ$ ,  $\Psi = 52^\circ$ .

(3) Asp<sup>19</sup> of staphylococcal nuclease: predicted  $\phi = -90 \pm 12^\circ$ ;  $\Psi = 8 \pm 11^\circ$ ; X-ray  $\phi = -156^\circ$ ,  $\Psi = -166^\circ$ .

Both for Asp<sup>159</sup> and Asp<sup>19</sup> there is no doubt regarding the similarity in backbone angles in solution and in the crystalline state, but TALOS fails to predict the unusual backbone angles of these residues. The user therefore should be aware that a small fraction of the TALOS predictions may be in error. However, as

shown below, for the vast majority of cases, the output of TALOS is highly accurate. When listing the rms differences between the predicted  $\phi/\Psi$  angles and those of the crystal structure, the small fraction of erroneous predictions are not included.

For ubiquitin, TALOS yields 53  $\phi/\Psi$  angle predictions (76% of its database residues) and the rms differences between the predicted  $\phi/\Psi$  angles and those of the crystal structure are  $12^\circ/9^\circ$ . Similarly, for cutinase  $\phi/\Psi$  predictions are made for 127 residues (69%, including 5 bad predictions, but excluding the disordered N-terminal tail), with rmsds of  $12^\circ/12^\circ$  relative to the crystallographically determined  $\phi$  and  $\Psi$  angles.

BPTI yielded the worst performance of all proteins tested. Only 32  $\phi/\Psi$  predictions (65%, 4 bad predictions) were made, which agree to within rmsds of  $16^\circ$  and  $17^\circ$  with the 1.1 Å crystal structure. Differences

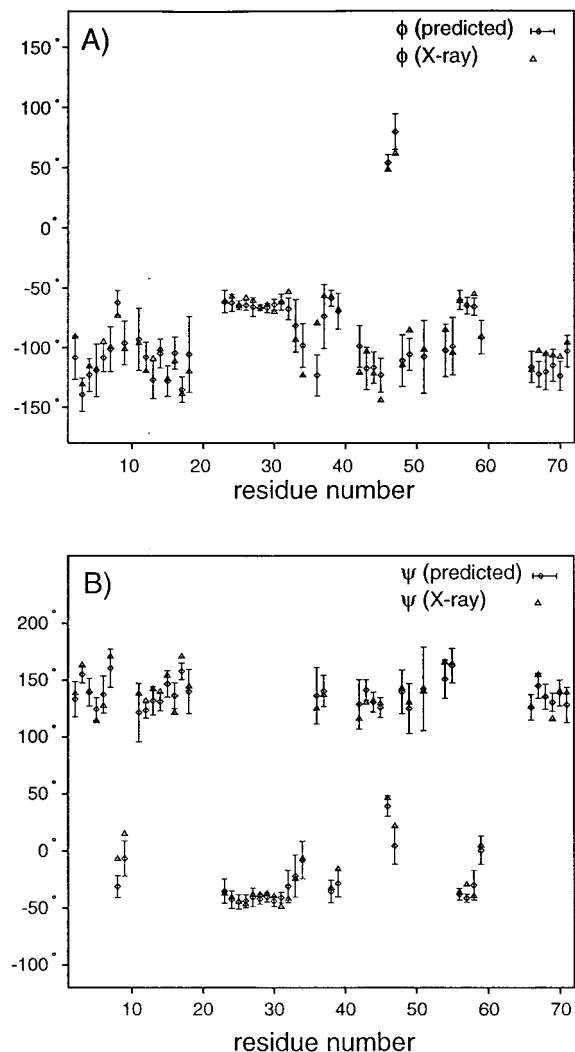


Figure 4. Predicted backbone angles (A)  $\phi$ , and (B)  $\psi$  of ubiquitin. The length of the error bars represents the standard deviation from the average of the dihedral angles of the 10 residues from the database having the highest chemical shift and sequence similarity with the query residues. Triangles correspond to the angles observed in the crystal structure.

relative to the solution structure (Berndt et al., 1992) are slightly larger ( $18/19^\circ$ ). For the same set of  $\phi$  and  $\psi$  angles, the rms differences between the average solution structure and crystal structure are  $14^\circ$  and  $12^\circ$ , respectively.

For human thioredoxin the NMR data have been derived for a mutant which differs from the sequence used for the crystal structure. The  $\phi$  angles predicted by TALOS are nevertheless in very good agreement with those of the crystal structure (Supplementary material), with 80 (78%)  $\phi/\psi$  predictions (rmsds of  $15^\circ$

and  $12^\circ$  from the X-ray structure, respectively), including one erroneous prediction. For reference, the rmsds relative to the solution structure for the same group of  $\phi$  and  $\psi$  angles are  $20^\circ$  and  $22^\circ$ , respectively. The pairwise rmsd between the crystal structure and solution structure angles is  $16^\circ$  ( $\phi$ ) and  $20^\circ$  ( $\psi$ ).

#### Use of TALOS output in structure calculation

The dihedral constraints for the backbone torsion angles obtained from TALOS are available immediately after completion of the resonance assignment and therefore can be used at the very early stages of structure calculation. It is, however, important to realize that a small fraction of the TALOS predictions is likely to be in error. Preliminary testing on the effect of inclusion of TALOS constraints in the calculation of a protein structure was carried out for ubiquitin.

Three sets of calculations were performed: (A) using only 273 NOEs, randomly taken from the total set of 2727 NOE cross peaks, peak-picked from 3D and 4D NOESY spectra (J.L. Marquardt, unpublished results); (B) additionally using TALOS- $\phi/\psi$  constraints for the 53 residues for which a (correct) prediction had been made; (C) as B, but deliberately introducing two serious errors in the  $\phi/\psi$  constraints by interchanging the TALOS-derived angles of Ala<sup>46</sup> (TALOS:  $\phi = 54 \pm 7^\circ$ ,  $\psi = 39 \pm 9^\circ$ ; X-ray:  $\phi = 48^\circ$ ,  $\psi = 46^\circ$ ) with those of Arg<sup>54</sup> (TALOS:  $\phi = -102 \pm 22^\circ$ ,  $\psi = 150 \pm 17^\circ$ ; X-ray:  $\phi = -85^\circ$ ,  $\psi = 165^\circ$ ). Starting from a fully extended strand and using an X-PLOR based simulated annealing protocol (Nilges et al., 1988), set A yielded convergence for 9 out of 30 calculated structures. The backbone rmsd (residues 2–70) from the average was  $1.52 \text{ \AA}$ , and the backbone rmsd displacement between the average of these NMR structures and the crystal structure was  $1.36 \text{ \AA}$ . For set B,  $\phi$ - and  $\psi$ -constraints were included as ‘harmonic-well’ potentials with zero energy over the range  $\phi_{\text{TALOS}} \pm \text{SD}$  and  $\psi_{\text{TALOS}} \pm \text{SD}$ , where SD is the standard deviation in the set of 10 (or 9) residues from which  $\phi_{\text{TALOS}}$  and  $\psi_{\text{TALOS}}$  were derived. Outside the well, the energy increased quadratically with  $200 \text{ kcal/rad}^2$ . With 13 out of 30 calculations converging, the yield was 50% higher than in the absence of TALOS constraints. Moreover, the rmsd from the average was also considerably lower ( $0.75 \text{ \AA}$ ), as was the difference relative to the X-ray structure ( $0.89 \text{ \AA}$ ). For set C, which includes the erroneous backbone constraints, convergence was worst (7 out of 30), but the rmsd values from the average ( $0.87 \text{ \AA}$ ) and between the averaged NMR and crystal structure ( $1.04 \text{ \AA}$ ) were

intermediate. The errors introduced in the NMR structure by the wrong TALOS constraints were highly localized.

Although preliminary and clearly incomplete, the above results for ubiquitin are quite encouraging. They suggest that a substantial improvement in quality of the structure can be obtained by including the TALOS-derived  $\phi/\Psi$ -restraints, particularly when the number of NOEs per residue is low. The introduction of two serious errors in the TALOS-derived torsion angle restraints decreases the quality of the structure, but it remains better than in the absence of the TALOS-derived constraints. Nevertheless, it is recommended that the constraints are used with care, keeping in mind that they may contain errors. Thus, if either a TALOS- or NOE-constraint (or both) is violated consistently during structure calculations, it is essential to recheck the quality of the constraint(s) involved. In this respect, an erroneous TALOS-derived restraint is no different from a wrongly assigned NOE connectivity.

### Concluding remarks

The approach described in this paper is the first to combine both chemical shift and residue type information for predicting backbone torsion angles. Also, instead of using the chemical shift information of only a single residue, it considers the chemical shifts and residue types of a string (of length 3, in the present case) to obtain this information. The weight of a particular secondary shift was adjusted by considering the width of its distribution over a narrow range of backbone torsion angles relative to the entire range of secondary chemical shifts in the database. The relative importance of the chemical shifts versus residue homology has been adjusted empirically to yield the most reliable predictions for proteins of known structure. Remarkably, the weighting factors for the center residue in the string of 3 residues in Table 1 is only slightly higher than for its two flanking residues, indicating that they are of comparable value when predicting a residue's  $\phi/\Psi$  angles. The contribution from the residue type homology to the similarity factor  $S$  is rather modest, typically about 25%. Nevertheless, reliability of TALOS predictions is considerably improved when including this residue type homology.

At the outset of developing this approach, we anticipated being able to obtain  $\chi_1$  angle predictions too. However, these  $\chi_1$  results so far appear insufficiently

reliable for general use. Three possible reasons for this are that (1) chemical shifts of the backbone nuclei are not sufficiently sensitive to  $\chi_1$ ; (2) in the crystal structures it is not possible to reliably and routinely separate residues with a single  $\chi_1$  conformation from those which undergo  $\chi_1$  rotameric averaging; and (3) there are practical difficulties in comparing  $\chi_1$  angles for residues with different types of side chains, i.e., a  $C^\beta$ -branched residue such as Thr with a non-branched residue. Although it may be feasible to develop criteria which yield useful TALOS  $\chi_1$  predictions, it is expected that it will be difficult to make predictions that are more reliable than those based on residue type and a residue's own backbone angles, as implemented by Kuszewski et al. (1997).

Our results indicate that concerted use of  $^{15}N$ ,  $^{13}C^\alpha$ ,  $^1H^\alpha$ ,  $^{13}C^\beta$  and  $^{13}C'$  chemical shifts of triplets of adjacent residues can be used to predict the backbone torsion angles for the majority of residues in assigned proteins. When using the crystal structure as the standard, the accuracy of the TALOS prediction appears to exceed that of even some of the best solution structures calculated on the basis of NOEs and J couplings. In principle, one could possibly argue that, as the angles in the database are all derived from crystal structures, one might expect the TALOS output to be closer to the crystal structure than to the solution structure. However, this argument is clearly invalid as it would require a systematic (as opposed to a random) difference between torsion angles in crystal structures and in solution. Second, when comparing the TALOS output for ubiquitin with a solution structure calculated by including a large number of  $^{13}C^\alpha$ - $^1H^\alpha$ ,  $^{13}C^\alpha$ - $^{13}C'$ ,  $^1H$ - $^{15}N$ ,  $^{13}C'$ - $^{15}N$  and  $^{13}C^\alpha$ - $^{13}C^\beta$  dipolar couplings (Tjandra and Bax, 1997; Marquardt et al., unpublished results) the agreement of the TALOS-predicted angles with the solution structure is actually better than with the crystal structure, with rmsds of 10° (solution) and 12° (X-ray) for  $\phi$  and 8° (solution) and 9° (X-ray) for  $\Psi$ . The rmsd between crystal structure and solution structure torsion angles is 7° for both  $\phi$  and  $\Psi$ .

The 3% fraction of TALOS predictions which are found to be in disagreement with the crystal structure includes residues which may adopt a different conformation in the solution and crystal structures (e.g.,  $Thr^{45}$  in cutinase, discussed above), although most of these regions where differences occur are excluded by the  $B$ -factor criterion (see Materials and methods). For most proteins used in our database, no high resolution solution structure is available, and it therefore was not possible to exclude these residues from the database.

A set of residues in the database for which the solution backbone angles differ strongly from those in the crystalline state does not increase the number of errors when TALOS is applied to a new protein. Instead, if their chemical shifts match those of the query triplet, they result in an outlier in the display of Figure 2. The same is true if a small fraction of residues in the database is wrongly assigned.

It also should be pointed out that a database approach such as the one described here tends to predict torsion angles that fall closer to the most commonly occupied regions of the Ramachandran map than the true value. This is a direct result of the fact that TALOS angles are derived from a set of triplets with the most similar chemical shifts: First, if the true backbone angles of a given center residue position it somewhere on the edge of the most populated region of the Ramachandran map, there statistically will be a larger number of 'hits' inside than outside the most populated region, simply because the density of residues is higher in the most populated region. This effect is visible in Figure 3B, for example, where for residues with X-ray  $\Psi$  angles in the  $-25^\circ$  to  $+25^\circ$  range the predicted  $\Psi$  angles are shifted in the direction of the  $\alpha$ -helical region of the Ramachandran map. Similarly, for residues with unusually large  $\Psi$  angles in the X-ray structure, the predicted values consistently are shifted slightly towards the more populated region near  $\Psi = 130^\circ$ . Second, in rare cases where residues are located far outside the populated region of the Ramachandran map (such as Asp<sup>19</sup> in staphylococcal nuclease), no other triplet with such unusual angles may be present in the database. If TALOS finds a cluster of triplets which accidentally match the shifts and residue types of the query triplet, it is likely that the torsion angles in this cluster fall in the highly populated region of the Ramachandran map. Both these types of problems will be alleviated when the database becomes larger.

It is important to realize that the TALOS-derived  $\phi/\Psi$ -values are empirical in nature. In a conservative approach, deviations between these  $\phi/\Psi$ -values and those in structures calculated on the basis of regular experimental restraints can be used for 'trouble-shooting' purposes. Alternatively, in cases where an insufficient number of regular experimental constraints is available, preliminary results on ubiquitin suggest that incorporation of the TALOS-derived  $\phi/\Psi$ -values can enhance structural quality considerably. Collecting a large number of NOEs can be particularly difficult in larger proteins, which require

extensive deuteration. It is expected that the use of TALOS-derived torsion angle restraints, when combined with one-bond dipolar couplings measured in dilute liquid crystalline media (Bax and Tjandra, 1997; Clore et al., 1998; Bewley et al., 1998; Hansen et al., 1998; Wang et al., 1998), will make it possible to obtain reliable backbone structures for such larger systems, even if only a limited number of NOEs is available.

## Software availability

The software, installation instructions and examples, are available upon request by electronic mail to [delaglio@speck.niddk.nih.gov](mailto:delaglio@speck.niddk.nih.gov). For further information see: <http://spin.niddk.nih.gov/bax/>.

## Acknowledgements

We thank Sharon Archer, Vladimir Basus, Rolf Boelens, Walter Chazin, Marius Clore, Bennett Farmer, Stephen Fesik, Kevin Gardner, Poul Hansen, Mitsuhiro Ikura, Marcel Ottiger, Jeanine Prompers, Sergio Scrofani, and Dennis Torchia for providing chemical shift assignments included in the database, and John Marquardt, Marcel Ottiger, and Jin-Shan Hu for useful discussions. Work by G. Cornilescu is in partial fulfillment for the Ph.D. degree at the University of Maryland, College Park, MD.

## References

- Ando, I., Saito, H., Tabeta, R., Shoji, A. and Ozaki, T. (1984) *Macromolecules*, **17**, 457–461.
- Archer, S.J., Vinson, V.K., Pollard, T.D. and Torchia, D.A. (1994) *FEBS Lett.*, **337**, 145–151.
- Bax, A. and Tjandra, N. (1997) *J. Biomol. NMR*, **10**, 289–292.
- Beger, D.B. and Bolton, P.H. (1997) *J. Biomol. NMR*, **10**, 129–142.
- Berndt, K.D., Guntert, P., Orbons, L.P. and Wüthrich, K. (1992) *J. Mol. Biol.*, **227**, 757–775.
- Betzel, C., Klupsch, S., Papendorf, G., Hastrup, S., Branner, S. and Wilson, K.S. (1992) *J. Mol. Biol.*, **223**, 427–445.
- Bewley, C.A., Gustafson, K.R., Boyd, M.R., Covell, D.G., Bax, A., Clore, G.M. and Gronenborn, A.M. (1998) *Nat. Struct. Biol.*, **5**, 571–578.
- Brünger, A.T. (1993) *XPLOR Manual*, Version 3.1, Yale University, New Haven, CT.
- Celda, B., Biamonti, C., Arnaud, M.J., Tejero, R. and Montelione, G.T. (1995) *J. Biomol. NMR*, **5**, 161–172.
- Chattopadhyaya, R., Meador, W.E., Means, A.R. and Quiocho, F.A. (1992) *J. Mol. Biol.*, **228**, 1177–1192.

- Clore, G.M., Bax, A., Driscoll, P.C., Wingfield, P. and Gronenborn, A. (1990) *Biochemistry*, **29**, 8172–8184.
- Clore, G.M., Starich, M.R. and Gronenborn, A.M. (1998) *J. Am. Chem. Soc.*, **120**, 10571–10572.
- Concha, N.O., Rasmussen, B.A., Bush, K. and Herzberg, O. (1996) *Structure*, **4**, 823–836.
- Copie, V., Battles, J.A., Schwab, J.M. and Torchia, D.A. (1996) *J. Biomol. NMR*, **7**, 335–340.
- Davis, J.H., Agard, D.A., Handel, T.M. and Basus, V.J. (1997) *J. Biomol. NMR*, **10**, 21–27.
- de Dios, A.C. and Oldfield, E. (1993) *J. Am. Chem. Soc.*, **116**, 5307–5314.
- de Dios, A.C., Pearson, J.G. and Oldfield, E. (1993) *Science*, **260**, 1491–1495.
- Delaglio, F., Grzesiek, S., Vuister, G., Zhu, G., Pfeifer, J. and Bax, A. (1995) *J. Biomol. NMR*, **6**, 277–293.
- Drakenberg, T., Hofman, T. and Chazin, W.J. (1989) *Biochemistry*, **28**, 5946–5954.
- Fedorov, A.A., Magnus, K.A., Graupe, M.H., Lattman, E.E., Pollard, T.D. and Almo, S.C. (1994) *Proc. Natl. Acad. Sci. USA*, **91**, 8636–8640.
- Fogh, R.H., Schipper, D., Boelens, R. and Kaptein, R. (1995) *J. Biomol. NMR*, **5**, 259–270.
- Fujinaga, M., Delbaere, L.T.J., Brayer, G.D. and James, M.N.G. (1985) *J. Mol. Biol.*, **184**, 479–502.
- Gardner, K.H., Zhang, X., Gehring, K. and Kay, L.E. (1998) *J. Am. Chem. Soc.*, **120**, 11738–11748.
- Gronenborn, A.M. and Clore, G.M. (1994) *J. Biomol. NMR*, **4**, 455–458.
- Gronwald, W., Boyko, R.F., Sönnichsen, F.D., Wishart, D.S. and Sykes, B.D. (1997) *J. Biomol. NMR*, **10**, 165–179.
- Hansen, P.E. (1991) *Biochemistry*, **30**, 10457–10466.
- Hansen, M.R., Rance, M. and Pardi, A. (1998) *J. Am. Chem. Soc.*, **120**, 11210–11211.
- Ikura, M., Kay, L.E. and Bax, A. (1990) *Biochemistry*, **29**, 4659–4667.
- Ikura, M., Kay, L.E., Krinks, M. and Bax, A. (1991) *Biochemistry*, **30**, 5498–5504.
- Ke, H.M., Zydowsky, L.D., Liu, J. and Walsh, C.T. (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 9483–9487.
- Kricheldorf, H.R. and Muller, D. (1983) *Macromolecules*, **16**, 615–623.
- Kumar, V. and Kannan, K.K. (1994) *J. Mol. Biol.*, **241**, 226–232.
- Kuntz, I.D., Kosen, P.A. and Craig, E.C. (1991) *J. Am. Chem. Soc.*, **113**, 1406–1408.
- Kuszewski, J., Qin, J., Gronenborn A.M. and Clore, G.M. (1995) *J. Magn. Reson.*, **B106**, 92–96.
- Kuszewski, J., Gronenborn, A.M. and Clore, G.M. (1997) *J. Magn. Reson.*, **125**, 171–177.
- Lam, P.Y.S., Jadhav, P.K., Eyerman, C.J., Hodge, C.N., Ru, Y., Bacheler, L.T., Meek, J.L., Otto, M.J., Rayner, M.M., Wong, Y.N., Chang, C.-H., Weber, P.C., Jackson, D.A., Sharpe, T.R. and Erickson-Viitanen, S. (1994) *Science*, **263**, 380–384.
- Leesong, M., Henderson, B.S., Gillig, J.R., Schwab, J.M. and Smith, J.L. (1996) *Structure*, **4**, 253–256.
- Loll, P.J. and Lattman, E.E. (1989) *Proteins Struct. Funct. Genet.*, **5**, 183–201.
- Longhi, S., Czjzek, M., Lamzin, V., Nicolas, A. and Cambillau, C. (1997) *J. Mol. Biol.*, **268**, 779–799.
- Lugrinbühl, P., Szyperski, T. and Wüthrich, K. (1995) *J. Magn. Reson.*, **109**, 229–233.
- Markley, J.L., Bax, A., Arata, Y., Hilbers, C.W., Kaptein, R., Sykes, B.D., Wright, P.E. and Wüthrich, K. (1998) *J. Biomol. NMR*, **12**, 1–23.
- Meador, W.E., Means, A.R. and Quiocho, F.A. (1992) *Science*, **257**, 1251–1255.
- Nilges, M., Gronenborn, A.M., Brünger, A.T. and Clore, G.M. (1988) *Protein Eng.*, **2**, 27–38.
- Ösapay, K. and Case, D.A. (1994) *J. Biomol. NMR*, **4**, 215–230.
- Ottiger, M., Zerbe, O., Güntert, P. and Wüthrich, K. (1997) *J. Mol. Biol.*, **272**, 64–81.
- Ousterhout, J.K. (1994) *Tcl and the Tk Toolkit*, Addison-Wesley, Reading, MA.
- Pardi, A., Wagner, G. and Wüthrich, K. (1983) *Eur. J. Biochem.*, **137**, 445–454.
- Pastore, A. and Saudek, V. (1990) *J. Magn. Reson.*, **90**, 165–176.
- Pearson, J.G., Wang, J., Markley, J.L., Le, H. and Oldfield, E. (1995) *J. Am. Chem. Soc.*, **117**, 8823–8829.
- Pelton, J.G., Torchia, D.A., Meadow, N.D., Wong, C. and Roseman, S. (1991) *Biochemistry*, **30**, 10043–10057.
- Prompers, J.J., Groenewegen, A., van Schaik, R.C., Pepermans, H.A.M. and Hilbers, C.W. (1997) *Protein Sci.*, **6**, 2375–2384.
- Qin, J., Clore, G.C. and Gronenborn, A.M. (1996) *Biochemistry*, **35**, 7–13.
- Redfield, C. and Robertson, J. (1991) *Proceedings of a NATO Advanced Research Workshop on Computational Aspects of the Study of Biological Macromolecules by NMR*, Plenum Press, New York, NY.
- Saito, H. (1986) *Magn. Reson. Chem.*, **24**, 835–852.
- Scrofani, S.D.B., Wright, P.E. and Dyson, J.H. (1998) *J. Biomol. NMR*, **12**, 201–202.
- Seavey, B.R., Farr, E.A., Westler, W.M. and Markley, L. (1991) *J. Biomol. NMR*, **1**, 217–236.
- Sethson, I., Edlund, U., Holak, T.A., Ross, A. and Johnson, B.-H. (1996) *J. Biomol. NMR*, **8**, 417–428.
- Sharff, A.J., Rodseth, L.E. and Quiocho, F.A. (1993) *Biochemistry*, **32**, 10553–10559.
- Spera, S. and Bax, A. (1991) *J. Am. Chem. Soc.*, **113**, 5491–5492.
- Svensson, L.A., Thulin, E. and Forsen, S. (1992) *J. Mol. Biol.*, **223**, 601–606.
- Veerapandian, B., Gilliland, G.L., Raag, R., Svensson, L.A., Masaui, Y., Hirai, Y. and Poulos, T.L. (1992) *Proteins Struct. Funct. Genet.*, **12**, 10–23.
- Vijay-Kumar, S., Bugg, C.E. and Cook, W.J. (1987) *J. Mol. Biol.*, **194**, 531–544.
- Vuister, G.W., Delaglio, F. and Bax, A. (1992) *J. Am. Chem. Soc.*, **114**, 9674–9675.
- Vuister, G.W., Delaglio, F. and Bax, A. (1993) *J. Biomol. NMR*, **3**, 67–80.
- Wang, A.C., Grzesiek, S., Tschudin, R., Lodi, P.J. and Bax, A. (1995) *J. Biomol. NMR*, **5**, 376–382.
- Wang, Y.-X., Marquardt, J.L., Wingfield, P., Stahl, S.J., Lee-Huang, S., Torchia, D.A. and Bax, A. (1998) *J. Am. Chem. Soc.*, **120**, 7385–7386.
- Weichsel, A., Gasdaska, J.R., Powis, G. and Montfort, W.R. (1996) *Structure*, **15**, 735–751.
- Williamson, M. (1990) *Biopolymers*, **29**, 1423–1431.
- Wishart, D.S., Sykes, B.D. and Richards, F.M. (1991) *J. Mol. Biol.*, **222**, 311–333.
- Wishart, D.S. and Sykes, B.D. (1994) *J. Biomol. NMR*, **4**, 171–180.
- Wishart, D.S., Colin, G.B., Holm, A., Hodges, R.S. and Sykes, B.D. (1995a) *J. Biomol. NMR*, **5**, 67–81.
- Wishart, D.S., Colin, G.B., Yao, J., Abildgaard, F., Dyson, H.J., Oldfield, E., Markley, J.L. and Sykes, B.D. (1995b) *J. Biomol. NMR*, **6**, 135–140.
- Wishart, D.S., Watson, M.S., Boyko, R.F. and Sykes, B.D. (1997) *J. Biomol. NMR*, **10**, 329–336.

- Wlodawer, A., Walter, J., Huber, R. and Sjolin, L. (1984) *J. Mol. Biol.*, **198**, 469–480.
- Worthy lake, D., Meadow, N.D., Roseman, S., Liao, D.-I., Herzberg, O. and Remington, S.J. (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 10382–10386.
- Yamazaki, T., Hinck, A.P., Wang, Y.-X., Nicholson, L.K., Torchia, D.A., Wingfield, P.T., Stahl, S.J., Kaufman, J.D., Chang, C.-H., Domaille, P.J. and Lam, P.Y.S. (1996) *Protein Sci.*, **5**, 495–506.