

The predictive accuracy of secondary chemical shifts is more affected by protein secondary structure than solvent environment

Marie-Laurence Tremblay · Aaron W. Banks ·
Jan K. Rainey

Received: 18 December 2009 / Accepted: 17 February 2010 / Published online: 6 March 2010
© Springer Science+Business Media B.V. 2010

Abstract Biomolecular NMR spectroscopy frequently employs estimates of protein secondary structure using secondary chemical shift ($\Delta\delta$) values, measured as the difference between experimental and random coil chemical shifts (RCCS). Most published random coil data have been determined in aqueous conditions, reasonable for non-membrane proteins, but potentially less relevant for membrane proteins. Two new RCCS sets are presented here, determined in dimethyl sulfoxide (DMSO) and chloroform:methanol:water (4:4:1 by volume) at 298 K. A web-based program, CS-CHEMeleon, has been implemented to determine the accuracy of secondary structure assessment by calculating and comparing $\Delta\delta$ values for various RCCS datasets. Using CS-CHEMeleon, $\Delta\delta$ predicted versus experimentally determined secondary structures were compared for large datasets of membrane and non-membrane proteins as a function of RCCS dataset, $\Delta\delta$ threshold, nucleus, localized parameter averaging and secondary structure type. Optimized $\Delta\delta$ thresholds are presented both for published and for the DMSO and chloroform:methanol:water derived RCCS tables. Despite obvious RCCS variations between datasets, prediction of secondary structure was consistently similar. Strikingly, predictive accuracy seems to be most dependent upon the

type of secondary structure, with helices being the most accurately predicted by $\Delta\delta$ values using five different RCCS tables. We suggest caution when using $\Delta\delta$ -based restraints in structure calculations as the underlying dataset may be biased. Comparative assessment of multiple RCCS datasets should be performed, and resulting $\Delta\delta$ -based restraints weighted appropriately relative to other experimental restraints.

Keywords Random coil chemical shifts · Nuclear magnetic resonance spectroscopy · Dielectric constant · Solvent effects · Secondary chemical shifts · Chemical shift index · Protein secondary structure

Introduction

Chemical shifts are often highly amino acid specific and can provide a strong indication of polypeptide secondary structure (Wishart et al. 1992; Wishart and Sykes 1994; Thanabal et al. 1994). Evaluation of chemical shifts typically relies on subtraction of representative random coil chemical shifts (RCCS) from an observed chemical shift in order to provide a secondary chemical shift ($\Delta\delta$) independent of the identity of the amino acid. The $\Delta\delta$ values of H^α , H^β , C^α , C^β and C' nuclei are particularly sensitive to secondary structure (Berjanskii and Wishart 2007) and are frequently used either directly as restraints in structure calculations (Kuszewski et al. 1996) or in calculation of chemical shift index (CSI) parameters used for identifying regions of secondary structure (Wishart et al. 1995a). In structural biology, chemical shifts are most often used to highlight regions of protein secondary structure and are implemented during structure calculation protocols in

Electronic supplementary material The online version of this article (doi:10.1007/s10858-010-9400-5) contains supplementary material, which is available to authorized users.

M.-L. Tremblay · A. W. Banks · J. K. Rainey (✉)
Department of Biochemistry & Molecular Biology, Dalhousie
University, Halifax, NS B3H 1X5, Canada
e-mail: jan.rainey@dal.ca

J. K. Rainey
Department of Chemistry, Dalhousie University, Halifax,
NS B3H 1X5, Canada

parallel with other distance, orientation or dihedral angle based restraints (Evans 1995).

Since the first report that chemical shifts are systematically affected by polypeptide secondary structure by Dalgarno et al. (1983), the relationship between chemical shifts and secondary structure has been extensively studied through empirical correlations (Szilagy and Jardetzky 1989; Zuiderweg et al. 1989; Williamson 1990; Williamson et al. 1992; Iwadata et al. 1999; Osapay and Case 1991; Spera and Bax 1991; Begler and Bolton 1997; Le and Oldfield 1994), determined through probabilistic methods (Lukin et al. 1997; Wang and Jardetzky 2002), quantum mechanically (Xu and Case 2001; Jiao et al. 1993; Case 1998, 2000), and derived experimentally (Wüthrich 1986; Thanabal et al. 1994; Merutka et al. 1995; Wishart et al. 1995a; Schwarzingler et al. 2001, 2000). In cases with complete chemical shift assignment, including ^{13}C and ^{15}N nuclei, bioinformatics approaches such as TALOS (Cornilescu et al. 1999), TALOS+ (Shen et al. 2009a) and PREDITOR (Berjanskii et al. 2006) use databases to produce comparable protein fragments, in some cases using homology considerations, to define ϕ and ψ dihedral angles. DANGLE extends these approaches through inclusion of Bayesian inference to improve sampling of less populated regions of Ramachandran space (Cheung et al. 2010). Algorithms which build upon these concepts for complete protein structural restraint generation by chemical shifts alone include CHESHIRE (Cavalli et al. 2007), CS-ROSETTA (Shen et al. 2008) and CS23D (Wishart et al. 2008), with CS-ROSETTA recently extended to support cases of incomplete chemical shift assignment (Shen et al. 2009b). Although these methods do not explicitly account for variation in dielectric constant (ϵ) or environment between protein interior and exterior, the chemical shift comparison databases inherently consist of structural fragments in a wide variety of environments.

With incomplete chemical shift assignments, i.e., if isotope labeling is infeasible, or if the direct comparability of the protein database employed for dihedral angle prediction to the protein studied is questionable, $\Delta\delta$ and CSI type approaches based on comparison to RCCS datasets are employed. Our ability to accurately assess secondary structure from RCCS is dependent on solvent conditions (Avbelj et al. 2004; Plaxco et al. 1997; Thanabal et al. 1994), effects of nearest neighbours (Schwarzingler et al. 2001; Wishart et al. 1995a), temperature (Merutka et al. 1995; Tonan and Ikawa 2003), and spectral referencing (Wishart et al. 1995b). In studies examining empirical relationships between chemical shifts and secondary structure, typically derived from datasets of globular proteins studied in aqueous conditions, it is difficult to untangle the effects of environmental differences between nuclei exposed to aqueous solution or buried in the

globular core of a protein versus the effects of polypeptide backbone structural differences. Furthermore, all currently available experimentally measured RCCS datasets have been determined in aqueous solution.

The solvent environment can perturb chemical shifts of solutes by a number of mechanisms, including effects of ϵ . Dielectric constant affects chemical shifts by altering the transmittance of the electrostatic field (Sondergaard et al. 2008). Other mechanisms by which solvent environment can affect chemical shifts include modification of preferred bond torsion angles, differences in hydrogen bonding with the solvent, van der Waals interactions, ring currents, and electric charge (Hass et al. 2008). Solvent-induced modification of protonation state at ionizable sites would also be expected to perturb chemical shifts of nearby nuclei (Kim et al. 2009). These other properties are solvent-specific and are not always correlated to the ϵ of the solvent. Given that chemical shifts are influenced by solvent (Shenderovich et al. 2001; Tonan and Ikawa 2003), there is a strong incentive to ensure that the environment used to determine RCCS matches that in which the polypeptide or protein is being studied. In particular, membrane proteins or fibrillar self-assembled proteins are found in liquid crystalline environments with substantially lower ϵ than water. Furthermore, the hydrophobic core of a globular protein is better represented as a region of low ϵ in comparison to water (GarciaMoreno et al. 1997).

In this paper, we present RCCS values determined in non-aqueous environments and examine the effects of perturbation of solvent environment on our ability to directly correlate $\Delta\delta$ to secondary structure. A set of 21 random coil peptides of sequence GGXAGG, where X is any of the 20 naturally occurring amino acids or the modified amino acid 4-hydroxyproline, was studied in two different media with ϵ lower than water. Dimethyl sulfoxide (DMSO; $\epsilon = 47.5$) was used as a mimic of the bilayer/water interface ($\epsilon \sim 40$) (Brockman 1994; Koehorst et al. 2008). A ternary solvent mixture (hereafter referred to as the trisolvant system; theoretical $\epsilon \sim 37.8$) composed of chloroform, methanol, and water in a 4:4:1 ratio (by volume) was also investigated due to its membrane mimetic properties (Slepkov et al. 2005). Notably, it is not clear which solvent perturbation effect(s) of those mentioned above would be most significant in the DMSO or the trisolvant system. However, since previous structural studies have successfully demonstrated both of these solvents to be reasonable membrane mimetics (reviewed in Rainey et al. (2006)), providing a good approximation of both the lower ϵ and decreased availability of H-bonding donors and acceptors in a membrane environment, RCCS values determined in these environments should also be applicable in a membrane environment. The python program CS-CHEMeleon, web-mounted at <http://structbio.biochem.dal.ca/jrainey/CSChem>, was implemented

to allow rapid analysis of sizeable sets of globular and membrane protein structures solved by NMR methods with published chemical shifts. This approach allowed comprehensive statistical comparison of the relative ability to predict protein secondary structure in membrane and non-membrane proteins using $\Delta\delta$ values derived from our sets versus three other sets of RCCS values, providing the first differential analysis (to our knowledge) between these classes of proteins.

Materials and methods

Materials

9-Fluorenylmethoxycarbonyl (Fmoc) protected amino acids, Rink Amide AM Sure Resin (0.65 mmol/g loading) and coupling reagents were obtained from AAPPTec (Louisville, KY), except for Fmoc protected ^{15}N -labelled Gly (Cambridge Isotope Laboratories, Andover, MA) and Ala (C/D/N Isotopes, Pointe-Claire, QC). N,N-dimethylformamide (sequencing grade) and acetonitrile (high performance liquid chromatography (HPLC) grade) were obtained from Fisher Scientific (Ottawa, ON). The deuterated solvents DMSO- d_6 and methanol- d_3 (CD_3OH) were acquired from C/D/N Isotopes while chloroform (CDCl_3) was obtained from Sigma–Aldrich (Oakville, ON). The chemical shift standard 2,2-dimethyl-2-silapentane-5-sulfonic acid (DSS) was obtained from Wilmad (Buena, NJ). All other chemicals were purchased as biotechnology, high performance liquid chromatography (HPLC) or reagent grade, as appropriate, from Sigma–Aldrich. All reagents and chemicals were used without further purification, unless otherwise specified. NMR samples were prepared in either 5 mm O.D. Wilmad 535-PP-7 or -8 tubes (DMSO and aqueous samples) or 535-TR-8 screw cap tubes (trisolvent mixture).

Peptide synthesis and purification

Peptides with sequence Ac-Gly-Gly-X-Ala-Gly-Gly-NH₂ (X being any of the 20 L-amino acids or the modified amino acid 4-hydroxyproline) were synthesized at ~ 0.2 mmol scale on a semi-automatic solid-phase peptide synthesizer (Endeavor 90, AAPPTec) on Rink AM resin. Protocols were as outlined in Langelan et al. (2009) with the exception that peptides were N-terminally acetylated with anhydrous acetic anhydride (5 eq. to resin) and the cleavage cocktails used were appropriate for the side-chain protecting groups in a given peptide (Guy and Fields 1997). Reverse phase HPLC (Beckman System Gold, Fullerton, CA) purification was performed using a C₁₈ column (5 μm particle, 120 Å pore size, 10 \times 250 mm AAPPTec Spirit)

at a flow rate of 3.0 mL/min. A linear water/acetonitrile (A/B) gradient was used from 98%A/2%B to 60%A/30%B over 25 min. Peptide identities and purities were confirmed by NMR spectroscopy.

Circular dichroism (CD) spectropolarimetry

Far-ultraviolet (far-UV) CD spectropolarimetry (J-810, Jasco, Easton, MD) was performed in 2,2,2-trifluoroethanol (Sigma, 99% non-deuterated NMR grade) for 5 peptides (peptide concentration determined by UV absorbance at 210 nm using a 1.0 cm path length quartz cuvette (Hellma, Müllheim, Germany) on a diode array spectrophotometer (Hewlett Packard, 8452A) for ellipticity normalization) with X = G, I, M, P, and V at 25°C (controlled with a NESLAB RTE-111 bath, Thermo Scientific, Newington, NH). Three repetitions (190–260 nm, 1 nm steps, 20 nm/min) were performed and averaged for all trials of each peptide in a 0.1 mm path length quartz cuvette (Hellma, Müllheim, Germany). All spectra were blank subtracted, treated with a weighted sliding-window average and converted to mean residue ellipticity $[\theta]$ as described previously (Langelan et al. 2009).

NMR spectroscopy

Experiments were performed at 298 K either on the Nuclear Magnetic Resonance Research Resource (NMR-3, Dalhousie University) 11.7 T Avance II spectrometer (Bruker Canada, Milton, ON) equipped with a 5 mm broadband observed (BBO) probe or the National Research Council Institute for Marine Biosciences (NRC-IMB, Halifax, NS) 16.4 T Avance III (Bruker Canada) spectrometer equipped with a 5 mm indirect detection TCI cryoprobe. Samples (10 mM peptide, 5 mM DSS, 600 μL) were prepared in DMSO- d_6 or CDCl_3 : CD_3OH : H_2O (4:4:1 by volume; mixture pH 5.5 ± 1). An aqueous sample of the Hyp peptide was prepared using the conditions of Wishart et al. (1995a). Spectra collected in DMSO are reported indirectly referenced to a value of 0 ppm for aqueous DSS using intermediate trimethylsilane (TMS) shifts (^1H (Hoffman 2006), ^{13}C (Hoffman 2003; Wishart et al. 1995b)). Shifts in the trisolvent system (and of the Hyp in aqueous solution) are reported relative to internal DSS at 0 ppm. The dielectric constant of the trisolvent system (ϵ_s) was estimated using a combination of equations published by Abraham et al. (1966) and Amirjahed and Blake (1975):

$$\epsilon_s = \sum \frac{\epsilon_n - 1}{\epsilon_n + 2} M_n \quad (1)$$

where ϵ_n is the dielectric constant of solvent n with mole fraction M_n and the sum is carried out over all components of the solvent mixture.

Table 1 Details of NMR experiments performed

Experiment	Bruker pulse program	Number of scans	Relaxation delay (s)	Sweepwidth (Hz)	Acquisition time (s)	Acquired points and increments
1D	zg30 or zgpgppr	64	2.00	7,002	3.28	32,768
DEPTQ135	deptqsp	4,000	1.00	33,784	0.74	50,670
^{13}C - ^1H HSQC	hspcetgpsisp.2	16	1.52	F1 33784 F2 4807	0.11	1,024 × 128
^{15}N - ^1H HSQC	hspcetfpf3gpsi	4	1.00	F1 3225 F2 8417	0.12	2,048 × 64
2D ^1H - ^1H TOCSY	dipsi2ph	8	1.50	F1 5122 F2 7183	0.28	4,096 × 32

Unless otherwise specified, experiments were performed at 11.7 T (details in Table 1). In cases where 1D ^1H NMR experiments were ambiguous, 1D nuclear Overhauser effect spectroscopy experiments (0.5 or 1 s mixing time) with 64 scans were performed by irradiation of the X amino acid H^α proton for identification of amide protons. A combination of distortionless enhancement by polarization transfer with modification for the detection of quaternary nuclei (DEPTQ-135; (Burger and Bigler 1998)) and 2D ^{13}C - ^1H sensitivity-enhanced heteronuclear single quantum coherence (HSQC; (Farrow et al. 1994)) allowed accurate ^{13}C shift assignment for all 21 peptides in all conditions. The ^{15}N -labelled Gly and Ala peptides were analyzed at 16.4 T. Sensitivity-enhanced ^{15}N - ^1H HSQCs (Farrow et al. 1994) were used to assign the labeled X -position amino acid ^{15}N and ^1H chemical shifts. 2D total correlation spectroscopy (TOCSY; 60 ms DIPSI-2 mixing time) was used to assign H^α and H^β for Ala and H^α for Gly. All 1D experiments were processed and analyzed using TopSpin 1.6 (Bruker) and 2D experiments were processed using NMRpipe (Delaglio et al. 1995) and analyzed using Sparky 3.115 (Goddard and Kneller 2008).

Comparative evaluation of random coil chemical shift tables

The DMSO and trisolvant system chemical shifts presented in this paper were compared to three sets of published aqueous chemical shifts, two derived experimentally and one by probability-based methods, for their ability to assess secondary structure in proteins. All 33 transmembrane (TM) protein NMR-STAR files currently in the BioMagResBank (BMRB: <http://www.bmrwisc.edu/> (Ulrich et al. 2008)) along with their matching structural data from the protein data bank of transmembrane proteins (PDB_TM: <http://pdbtm.enzim.hu/> (Tusnady et al. 2005)) were acquired (listed in the Supplementary Material). Chemical shifts and structures for a randomly selected set of 107 non-membrane proteins whose structures were determined in aqueous

conditions (AQ) were acquired from the BMRB and the Protein Data Bank (www.pdb.org (Berman et al. 2000)), respectively. All PDB files with less than 10 models were eliminated; 2 cases (i.e., 4 structures out of 107) of identical proteins having differing structures and chemical shift were retained (PDB entries 2KAX and 2KAY; 1YZA and 1YZC), and an instance of protein structures for a pair of isoforms having different lengths and C-terminal extensions was retained (PDB entries 2YSE and 2ZAJ). Beyond these 6 cases, the remaining 101 proteins were distinct based on pairwise comparison of sequence identity, with <75% identity over 20 amino acid stretches (excepting a pair of matrix metalloproteinases, PDB entries 1YCM and 2JNP, which have <64% pairwise identity over their entire lengths). Statistical examination was performed on concatenated datasets containing all proteins and structural models of each class (TM or AQ) rather than on individual NMR-STAR or PDB files. NMR-STAR files not referenced to DSS, with the assumption that those with unspecified reference were referenced to DSS, were indirectly referenced to DSS when the standard was trimethylsilyl-2,2,3,3-tetradeuteriopropionic acid (TSP), TMS, or H_2O for ^1H and TSP, TMS, or dioxane for ^{13}C (Hoffman 2006, 2003; Wishart et al. 1995b).

For the purpose of this study, the structure at a given amino acid residue “ i ” was determined from the average ϕ and ψ dihedral angles of the NMR-derived structural ensemble. For cases of localized averaging, a 5-residue sliding window average (residues $i - 2$ to $i + 2$) averaged over all ensemble members was employed. The consensus was deemed the “true” structure of the peptide or protein. Backbone dihedral angles were calculated and the regions defining various secondary structures were geometrically derived from those of Lovell et al. (2003) (exact boundaries detailed in the Supplementary Material). The reliability of local structural averaging was verified against 10 randomly selected proteins from the AQ dataset by qualitative comparison of our secondary structure output (helix, sheet, or coil) to that of PROMOTIF v1.0 (Hutchinson and Thornton 1996).

The chemical shift based secondary structure prediction (i.e., helix, coil or sheet) for each residue was determined by comparison of the magnitude and sign of the $\Delta\delta$ relative to a specified threshold for a given nucleus, with $\Delta\delta$ values calculated including the correction factors of Schwarzingger et al. for nearest neighbour effects (Schwarzingger et al. 2001). Optimized thresholds for nuclei were determined iteratively for each RCCS table, with calculation of agreement using thresholds of -1 ppm to $+1$ ppm (0.01 ppm interval) for H^α and of -2 to $+2$ (0.1 ppm interval) for C^α and C^β to obtain the best agreement between PDB file based dihedral angles and $\Delta\delta$ for the full AQ dataset (the TM dataset contained too few amino acids). Consensus $\Delta\delta$ -based predictions were defined as follows, based on the nuclei out of H^α , C^α , and C^β that are reported in an NMR-STAR file. If at least two of these nuclei have chemical shifts provided, the consensus prediction required agreement of at least 2 $\Delta\delta$ -based threshold tests. If there was no consensus, then the amino acid was designated as part of a coiled structure. Local averaging, when employed, was carried out identically to the local structural averaging using a 5-residue sliding window. The numbers of correct and incorrect predictions were compared for each threshold during iteration and the ‘correct-incorrect’ predictions were normalized to a percentage. The optimal threshold for that atom’s structure type was at the highest normalization value. The standard thresholds of 0.1 ppm for 1H and 0.7 ppm for ^{13}C introduced by Wishart et al. (1992) were used alongside the optimized thresholds determined for each RCCS dataset during analysis. The accuracy of RCCS assessment of secondary structure was expressed as a total percentage of correctly assessed structure for both the AQ and TM datasets, subdivided by RCCS set, atom type and secondary structure threshold. Glycine and proline were not used for comparison since they have different Ramachandran angle preferences from the other amino acids in the same secondary structures (Lovell et al. 2003) and also have a greatly elevated $\Delta\delta$ threshold (Wishart et al. 1992).

CS-CHEMeleon implementation

For data analysis, a python 2.5 program named CS-CHEMeleon was written. This has been configured to run from a web-based graphical user interface (freely available at <http://structbio.biochem.dal.ca/jrainey/CSChem>). Using a chemical shift file uploaded in NMR-STAR format (Ulrich et al. 2008), $\Delta\delta$ values are calculated directly using the RCCS dataset(s) specified by the user. By default, two experimentally determined aqueous chemical shift datasets (Schwarzingger et al. 2000; Wishart et al. 1995a), one aqueous probability-based RCCS table (Wang and Jardtzyk 2002), and the DMSO and trisolvant tables

presented herein can be used. Any desired alternative RCCS table may also be uploaded and used. $\Delta\delta$ values may be determined with or without accounting for nearest neighbours, using the correction factors published by Schwarzingger et al. (2001). Evaluation and comparison of $\Delta\delta$ values may be performed graphically within the web browser and/or $\Delta\delta$ values may be downloaded in ASCII format for offline analysis. The user also has the option to only assess secondary structure for residues inside or outside the membrane as defined in the associated extensible markup language PDB_TM file when investigating TM proteins.

CS-CHEMeleon also calculates ϕ and ψ dihedral angles upon demand for an uploaded NMR structural ensemble (or single structure) in PDB file format, allowing comparison between the secondary structure given by the region of the Ramachandran plot (Lovell et al. 2003) and the predicted secondary structure from a given RCCS table using the structural thresholds of Wishart et al. (1995a) or the optimized iterated thresholds presented here. The user can also make use of local sliding window structural averaging (in either the NMR-STAR, PDB or both files) if the assessment yields poor correlation to the PDB structure. The number of residues averaged (recommended 5) and the minimum proportion of agreement for a consensus definition over the window (recommended 0.51) are both defined by the user.

Results and discussion

Verification of peptide random coil character

The established random coil peptide series of Wishart et al. (1995a), with sequence Ac-GGXAGG-NH₂, was used herein. In order to ensure that random coil character was maintained in organic solvent conditions, far-UV CD spectropolarimetry was performed on a sample set of 5 peptides containing a variety of *X* amino acids (Fig. 1). Because DMSO is not a suitable solvent for CD due to absorbance in the far-UV region and because of incomplete solubility of all 21 peptides in the trisolvant system, the α -helix inducing solvent trifluoroethanol (TFE) (Merutka et al. 1995) was used. All five peptides, including some with *X* residues having high helical propensity in aqueous (Blaber et al. 1993) or membrane-mimetic environments (Li and Deber 1994), show a strong negative band at ~ 198 nm characteristic of a random coil (Greenfield 2004) versus the positive band expected of a helix or sheet structure. Weak negative ellipticity in the ~ 210 – 230 nm region is also observed in most of these peptides, with varying strengths. The CD banding in this region is not indicative of sheet or helix (Greenfield 2004), particularly

in the absence of a positive band at ~ 200 nm. Also informatively, the GGPAGG spectrum (most likely of all of the peptides to have polyproline-II character) contains no band structure indicative of polyproline-II character

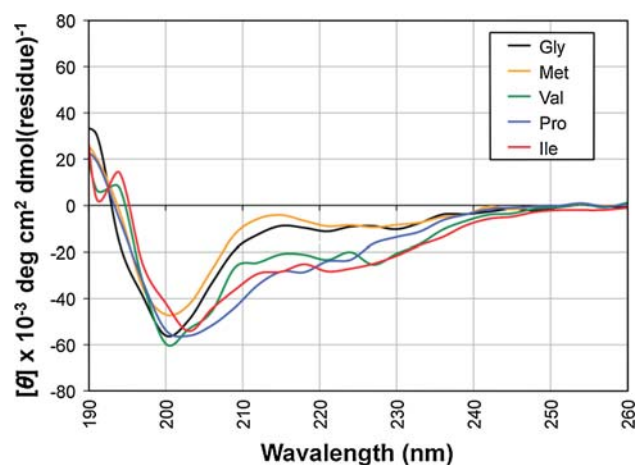


Fig. 1 Far-ultraviolet circular dichroism spectra of five random coil peptides (sequence Acetyl-GGXAGG-NH₂, with X given in legend) in 2,2,2-trifluoroethanol. The shown spectra are averages of three blank subtracted trials with ellipticity normalized to relative concentrations using UV-Vis absorbance at 210 nm, with a weighted 3 nm sliding-window average applied (detailed in “Methods”)

(Chellgren and Creamer 2004; Rath et al. 2005). We cannot explicitly rule out the weak negative ellipticity at 210–230 nm being caused by increased favourability of intramolecular or intermolecular interactions within or between peptides in TFE, relative to aqueous conditions. However, since all peptides examined had predominantly random coil CD characteristics even in an α -helical structure-inducing medium, we feel that the Ac-GGXAGG-NH₂ sequence, which was random coil in aqueous medium (Wishart et al. 1995a), should serve as a valid random coil model system in the non-structure inducing DMSO and trisolvant system environments.

Random coil chemical shifts

The complete set of RCCS measured in DMSO is reported in Tables 2 and 3 and in the trisolvant system in Tables 4 and 5. Note that only 10 of the 21 random coil peptides dissolved in the trisolvant system with no discernable pattern in solubility. For comparison, the chemical shifts for Hyp measured using identical conditions to Wishart et al. (1995a) are: H ^{α} 4.53; H ^{β} 2.35 and 2.06; H ^{γ} 4.62; H ^{δ} 3.79 and 3.57; C ^{α} 61.56; C ^{β} 39.76; C ^{γ} 72.5; and, C ^{δ} 57.3. Although DSS was incorporated as an internal standard in the DMSO samples, DSS and DMSO interact. Chemical

Table 2 ¹H random coil chemical shifts for peptides of sequence GGXAGG measured in dimethyl sulfoxide

X residue	NH	H ^{α}	H ^{β}	Others
Ala	8.21	4.47	1.43	
Arg	8.32	4.48	1.91, 1.74	γ CH ₂ 1.51, 1.51 δ CH ₂ 3.1
Asn	8.40	4.74	2.78, 2.65	γ NH ₂ 7.23, 7.67
Asp	8.41	4.78	2.93, 2.72	γ OH 12.6
Cys (red)	8.44	4.60	3.00, 2.92	SH 2.32
Gln	8.30	4.42	2.09, 1.94	γ CH ₂ : 2.30 δ NH ₂ : 6.97, 7.45
Glu	8.28	4.46	2.12, 1.95	γ CH ₂ 2.45 δ OH: 12.3
Gly	8.25	3.90		
His	8.43	4.80	3.31, 3.17	2CH 8.96 4CH 7.38
Hyp (<i>cis</i>)		4.74	2.39, 2.21	γ CH 4.45 δ CH ₂ 3.87, 3.57
Hyp (<i>trans</i>)		4.54	2.23, 2.09	γ CH 4.54 δ CH ₂ 4.15, 3.68
Ile	8.06	4.37	1.92	γ CH ₂ 1.62, 1.27 γ CH ₃ 0.85 δ CH ₂ 0.82
Leu	8.15	4.49	1.67, 1.67	γ CH ₂ 1.80 δ CH ₃ 1.09, 1.04
Lys	8.34	4.46	1.89, 1.73	γ CH ₂ 1.51 δ CH ₂ 1.72 ϵ CH ₂ 2.96
Met	8.29	4.54	2.13, 2.00	γ CH ₂ 2.66, 2.62 δ CH ₃ : 2.24
Phe	8.41	4.68	3.24, 2.96	2,6CH 7.38, 3,5CH 7.43 4CH 7.37
Pro (<i>cis</i>)		4.68	2.39, 2.21	γ CH ₂ 1.97 δ CH ₂ 3.63, 3.60
Pro (<i>trans</i>)		4.50	2.25, 2.07	γ CH ₂ 2.09 δ CH ₂ 3.75, 3.69
Ser	8.37	4.50	3.81, 3.78	
Thr	8.00	4.41	4.21	γ CH ₃ 1.45
Trp	8.37	4.69	3.35, 3.13	2CH 7.35 4H 7.79, 5H 7.24, 6CH 7.16 7H 7.50
Tyr	8.38	4.60	3.12, 2.86	2,6CH 6.84 3,5CH 7.23
Val	8.37	4.36	2.18	γ CH ₂ 1.06, 1.02

Table 3 ^{13}C random coil chemical shifts for peptides of sequence GGXAGG measured in dimethyl sulfoxide

X residue	C^α	C^β	Others
Ala	51.4	20.9	
Arg	55.2	32.1	γCH_2 28.0 δCH_2 43.6 ϵC 159.8
Asn	52.8	40.2	
Asp	52.6	39.1	
Cys (red)	58.1	29.3	
Gln	55.5	29.6	γC : 33.4
Glu	53.9	30.2	γCH_2 32.1
Gly	45.1		
His	54.5	30.1	1C 132.3 2CH 136.4 4CH 120.2
Hyp (<i>cis</i>)	60.8	40.7	γCH 70.1 δCH_2 58.4
Hyp (<i>trans</i>)	62.0	43.3	γCH 70.8 δCH_2 57.9
Ile	59.9	39.7	γCH_2 26.6 γCH_3 18.4 δCH_2 14.2
Leu	54.1	43.6	γCH_2 27.2 δCH_3 26.2, 24.6
Lys	55.3	34.3	γCH_2 25.2 δCH_2 29.7 ϵCH_2 41.9
Met	54.9	34.8	γCH_2 32.6 δCH_3 17.7
Phe	57.1	40.4	1C 140.9 2,6CH 132.3 3,5CH 131.2 4CH 129.4
Pro (<i>cis</i>)	61.7	34.8	γCH_2 25.2 δCH_2 49.9
Pro (<i>trans</i>)	62.9	32.2	γCH_2 27.4 δCH_2 49.1
Ser	58.2	64.7	
Thr	61.3	69.6	γCH_3 22.8
Trp	58.4	30.6	2CH 126.9 3C 113.0 4CH 121.5 5CH 124.0 6CH 121.3 7CH 114.4 8C 139.2 9C 130.4
Tyr	57.5	39.6	1C 130.9 3,5CH 118.0 2,6CH 133.2 4C 158.9
Val	60.7	33.5	γCH_3 21.2, 20.7

Table 4 ^1H random coil chemical shifts for peptides of sequence GGXAGG measured in methanol:chloroform:water (4:4:1 by volume)

X residue	NH	H^α	H^β	Others
Arg	8.30	4.34	1.89, 1.76	γCH_2 1.51, 1.51 δCH_2 3.1
Gly	8.26	3.86		
His	8.46	4.70	3.31, 3.16	2CH 8.56 4CH 7.30
Hyp (<i>trans</i>)		4.50	2.34, 2.05	γCH 3.33 δCH_2 3.80, 3.56
Leu	8.13	4.30	1.64, 1.65	γCH_2 1.68 δCH_3 0.96, 0.91
Phe	8.13	4.56	3.19, 3.01	2,6CH 7.26 3,5CH 7.30 4CH 7.23
Pro (<i>trans</i>)		4.39	2.27, 2.01	γCH_2 2.04 δCH_2 3.73, 3.59
Thr	8.24	4.32	4.27	γCH_3 1.27
Tyr	8.13	4.49	3.06, 2.93	2,6CH 6.76 3,5CH 7.09
Val	8.24	4.10	2.14	γCH_2 0.98, 0.97

Table 5 ^{13}C random coil chemical shifts for peptides of sequence GGXAGG measured in methanol:chloroform:water (4:4:1 by volume)

X residue	C^α	C^β	Others
Arg	53.6	30.6	γCH_2 26.4 δCH_2 42.0 ϵC 158.2
Gly	45.0		
His	54.6	29.0	1C 131.3 2CH 135.9 4CH 120.2
Hyp (<i>trans</i>)	62.2	39.9	γCH 72.0 δCH_2 57.1
Leu	52.6	42.1	γCH_2 25.6 δCH_3 24.7, 23.0
Phe	57.9	39.3	1C 138.9 2,6CH 131.4 3,5CH 130.8 4CH 129.2
Pro (<i>trans</i>)	63.5	31.7	γCH_2 24.2 δCH_2 49.2
Thr	59.8	68.1	γCH_3 21.2
Tyr	56.0	38.1	1C 129.4 3,5CH 116.5 2,6CH 131.7 4C 157.3
Val	57.3	32.4	γCH_3 20.1

shifts in these samples were therefore indirectly referenced to DSS in water using published reference values (Hoffman 2003, 2006; Wishart et al. 1995b) to provide direct comparability for experimental data acquired in aqueous solution using the accepted biomolecular chemical shift standard of DSS (Wishart et al. 1995a). In contrast, the shifts reported in the trisolvant system were internally

referenced to DSS since there is no straightforward way to indirectly reference in a ternary solvent system. Phase separation is noticeable in this solvent mixture (Slepkov et al. 2005), so it is likely that DSS referencing is very similar to that of DSS in water since DSS would be most soluble in water-rich components of the mixture. However,

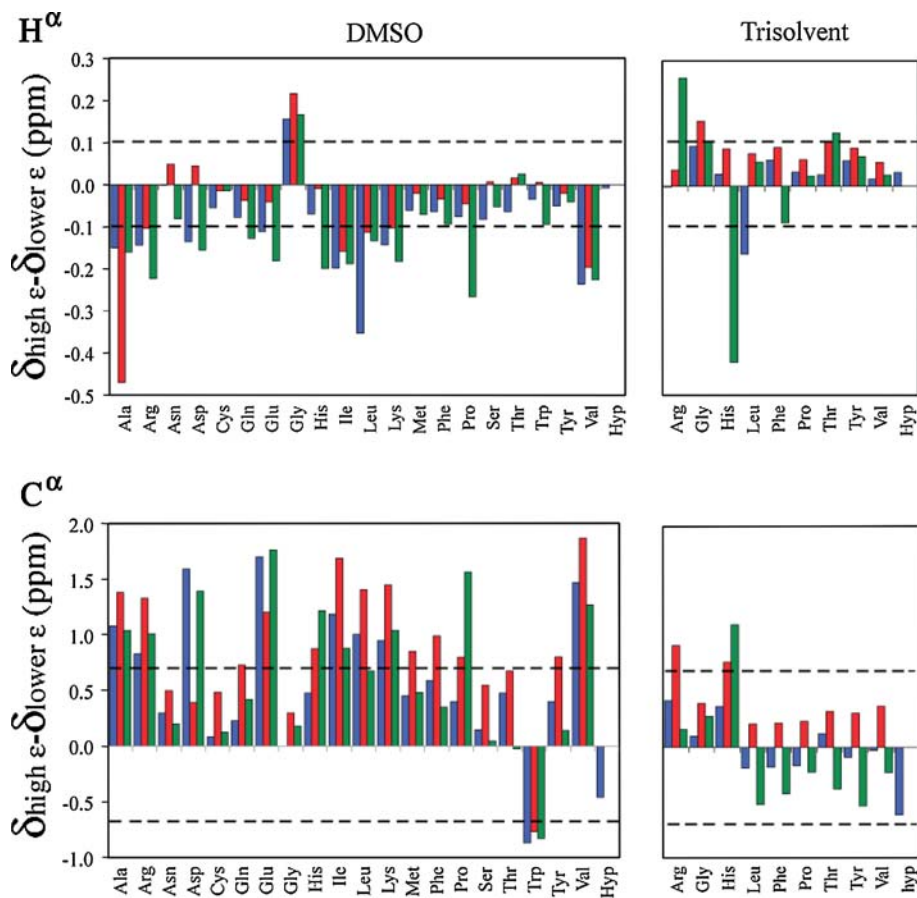
this uncertainty in referencing should be taken into account when employing the trisolvent-derived shifts. Ideally, RCCS would also be determined in a solvent of much lower ϵ for direct comparability to phospholipid tail-group regions or in the core of a globular protein. However, examination of a variety of solvents with lower ϵ (~ 4 – 20) demonstrated uniformly extremely poor solubilization of the random coil peptides. Designing a more hydrophobic, but still random coil, peptide would likely be required in order to obtain RCCS in a low ϵ medium.

The Ac-GGXAGG-NH₂ peptide series allows determination of the effect of solvent environment in terms of decreased ϵ and the other perturbation factors discussed in the introduction upon RCCS. These RCCS are most directly comparable to the experimental aqueous RCCS dataset of Wishart et al. (1995a), since the same peptide series was employed. Comparison was also performed to the experimental aqueous RCCS dataset of Schwarzingger et al. (2000) and to the statistically derived amino acid chemical shift dataset of Wang and Jardetzky (2002). RCCS perturbations are evident to different degrees for different amino acids and, in some instances, for different RCCS comparison sets (Fig. 2). Perturbations relative to

aqueous or statistically derived RCCS may be presumed to be arising from decreased ϵ , from changes to favoured dihedral angle ranges within a given random coil peptide from other non- ϵ derived properties of the solvent and, in the case of ionizable residues or H-bonding donor/acceptor side-chains, from protonation or H-bonding state. The aliphatic residues, for example, tend to be strongly perturbed in DMSO. Ionizable or H-bonding donor/acceptor side-chains, such as Asn and Gln, are also strongly perturbed in DMSO. As would be expected (Quirt et al. 1974), C $^{\alpha}$ is also strongly affected in the acidic residues, which are clearly protonated on the side-chain carboxylic acid in DMSO (Table 2) but typically deprotonated in aqueous conditions. H $^{\alpha}$ and C $^{\alpha}$ in DMSO generally (but not uniformly) experience opposite trends relative to aqueous RCCS, with H $^{\alpha}$ being deshielded while C $^{\alpha}$ is shielded. For both nuclei, more than half of the amino acids' random coil values change by more than their respective structural thresholds defined by Wishart et al. (1995a), which should therefore be defined as a significant change.

In comparison, the chemical shifts determined in the trisolvent system show different trends from those observed with DMSO (Fig. 2). Although only 10 amino

Fig. 2 Chemical shift differences between random coil H $^{\alpha}$ and C $^{\alpha}$ nuclei determined in high (aqueous) and intermediate (DMSO or trisolvent) dielectric environments. Aqueous random coil chemical shifts are from Wishart et al. (1995a) (blue), Schwarzingger et al. (2000) (red), and Wang and Jardetzky (2002) (green) except for the modified amino acid 4-hydroxyproline (Hyp), which is tabulated herein in aqueous conditions duplicating those of Wishart et al. (1995a)



acids can be compared, a change in solvent from aqueous conditions to the trisolvent system does cause perturbations in both H^α and C^α chemical shifts. In comparison to the DMSO environment, a lower proportion of chemical shifts are perturbed by more than the standard thresholds in the trisolvent system and, as a set, they do not show any trend in shielding and deshielding although the theoretical ϵ is lower than that of DMSO. This implies that effects of ϵ alone are not entirely responsible for the generally larger chemical shift perturbations observed in DMSO. Since some degree of phase separation is obvious in the trisolvent system (Slepkov et al. 2005), it is possible that preferential peptide solvation in aqueous-rich phases is giving rise to a decreased difference from RCCS derived in aqueous conditions. This hypothesis, however, is at odds with the insolubility of 11/21 of the random coil peptides in the trisolvent system.

Statistical analysis of NMR structure and chemical shift datasets

The accuracy of $\Delta\delta$ -based secondary structure prediction, derived using RCCS in aqueous vs. DMSO and trisolvent solutions, in both AQ and TM proteins datasets was assessed by the python program CS-CHEMeleon. A $\Delta\delta$ based prediction was assigned either using a single nucleus type or as the consensus (i.e., agreement by at least 2) of the H^α , C^α , and C^β $\Delta\delta$ values considered relative to the threshold in question. Since RCCS differ with solvent (Fig. 2), and since both DMSO and the trisolvent system are established membrane-mimetics, it was expected that the secondary structure of proteins in the TM dataset would have a higher agreement to the PBD structure when $\Delta\delta$ was calculated with RCCS in DMSO or the trisolvent system versus in aqueous conditions (and vice versa for the AQ

Table 6 Agreement between secondary chemical shift (requiring consensus of 2 of 3 of H^α , C^α , and C^β) based prediction of helix, sheet and coiled secondary structure in comparison to the consensus

structure determined by average ϕ and ψ dihedral angles at that position in the ensemble of NMR structures

Solvent	Type of 2° structure	Non-membrane proteins				Membrane proteins			
		No local avg.		Local avg.		No local avg.		Local avg.	
		Wishart	Iteration	Wishart	Iteration	Wishart	Iteration	Wishart	Iteration
DMSO	All	62.8	65.4	68.3	67.8	85.7	86.3	86.3	87.5
	H	86.0	84.2	92.7	87.6	97.3	95.7	98.6	98.3
	S	45.7	47.2	41.6	38.1	75.0 ^a	80.9 ^a	73.3 ^a	77.4
	C	27.8	33.0	44.3	58.8	21.7 ^a	8.7 ^a	30.0 ^a	25.0 ^a
Schwarzinger	All	63.4	68.0	69.7	71.0	87.5	87.9	91.2	88.1
	H	75.6	84.7	81.0	92.1	93.0	98.7	96.6	99.7
	S	57.6	56.1	57.2	54.1	87.2	80.3	87.2	77.9
	C	36.2	28.4	58.6	42.0	17.4 ^a	8.7 ^a	50.0 ^a	15.0 ^a
Wishart	All	63.3	67.7	69.5	71.9	87.5	86.1	92.4	89.8
	H	72.8	83.2	77.9	89.7	93.0	96.3	96.6	99.3
	S	57.6	56.1	57.2	54.1	86.2	81.4	88.2	83.6
	C	40.8	31.1	61.3	49.4	4.3 ^a	4.3 ^a	40.0 ^a	20.0 ^a
Wang	All	62.3	66.9	67.6	71.4	87.5	86.1	92.4	89.8
	H	69.9	81.8	73.3	88.8	92.3	94.7	95.9	97.3
	S	61.1	54.6	62.4	52.0	88.8	81.9	91.8	84.1
	C	38.2	36.2	60.3	54.8	13.0 ^a	8.7 ^a	45.0 ^a	35.0 ^a
Trisolvent	All	63.7	66.0	61.6	62.7	87.9	76.7	87.1	83.0
	H	71.1	78.2	66.6	74.6	93.9	89.7	95.2	95.2
	S	64.2	59.5	56.8	45.9	86.0	68.6	78.0	71.6
	C	37.7	36.8	57.2	61.9	20.0 ^a	13.5 ^a	66.7 ^a	22.2 ^a

Secondary chemical shift based predictions are compared using the “original” threshold values and the optimized values determined herein (Table 7) and using localized averaging (“Local avg.”, the 5-residue consensus both of secondary chemical shift predictions and dihedral angle based structure) versus consideration of secondary chemical shifts and dihedral angles at each residue in isolation (“No avg.”). Results are presented for datasets consisting of 109 non-membrane proteins and 19 membrane proteins, split into overall agreement (All) and agreement by secondary structure type of helix (H), sheet (S) and coil (C) The total number of amino acids compared are listed in Table 8

^a Sample size with less than 30 amino acids

Fig. 3 Relative predictive accuracy of secondary structure by secondary chemical shift for the indicated nucleus type for the non-membrane (AQ) and membrane (TM) protein datasets (dataset details in Table 8). The percentages of correctly predicted structure per amino acid using iteratively optimized secondary chemical shift thresholds (Table 7) were calculated as a function of nucleus (H^α , C^α , C^β , and consensus ≥ 2 of the three atoms), random coil chemical shift, and secondary structure type (helix (H), sheet (S), and coil (C)). Localized averaging over 5 residues was applied for both the secondary structure derived by ϕ and ψ dihedrals and for secondary chemical shift predictions

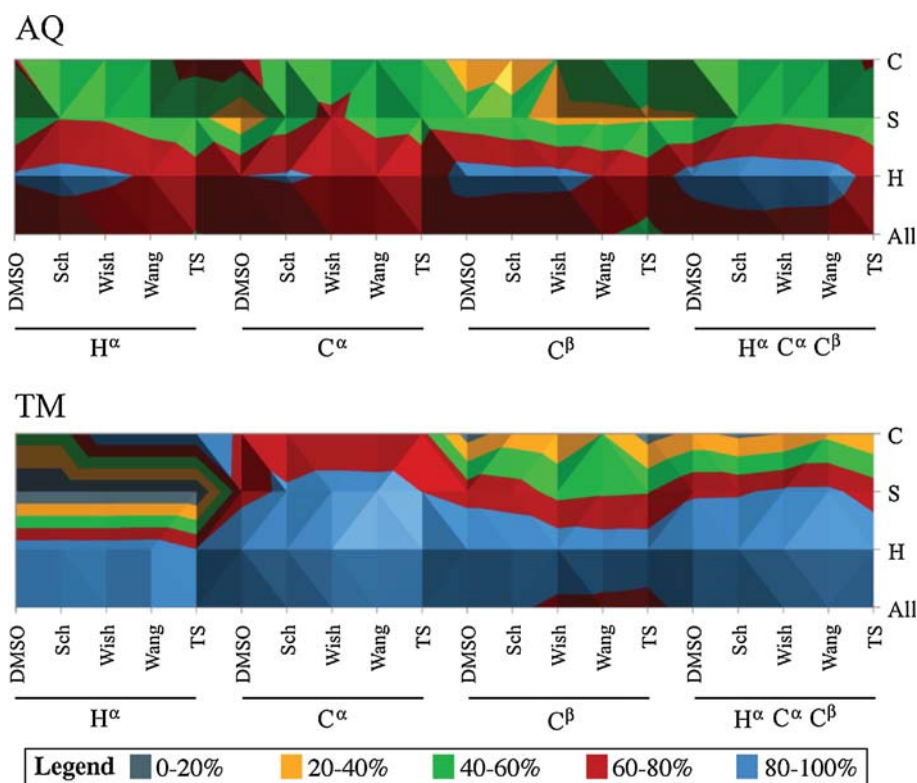


Table 7 Iteratively optimized thresholds for secondary chemical shift based prediction of helix, sheet, and coiled structures for the RCCS presented herein (Tables 2, 3, 4, 5) and for three published random coil sets: Schwarzinger et al. (2000) (Schwar.), Wishart et al. (1995a) (Wishart), Wang and Jardetzky (2002) (Wang)

	Original	DMSO	Schwar.	Wishart	Wang	Trisolvent
H^α -Helix	<-0.1	<-0.16	<-0.10	<-0.06	<-0.02	<-0.06
H^α -Sheet	>0.1	>0.22	>0.08	>0.16	>0.16	>0.24
C^α -Helix	>0.7	>1.30	>0.20	>0.50	>0.60	>0.80
C^α -Sheet	<-0.7	<-0.10	<-0.90	<-0.40	<-0.60	<-0.40
C^β -Helix	<-0.7	<-0.30	<0.20	<0.20	<0.20	<0.20
C^β -sheet	>0.7	>0.20	>1.00	>1.30	>1.60	>1.20

Coils are defined by secondary chemical shifts between the helix and sheet threshold values. The ‘original’ structural thresholds were those defined by Wishart et al. (1995a) and included for comparison. A set of 109 proteins with published structure and chemical shifts (the AQ dataset in Table 8) were used for optimization. Full optimization results are presented in the Supplementary Material

dataset). Using the original thresholds for defining helix, coil or sheet defined by Wishart et al. (1995a), the most obvious observation is the incredible accuracy of $\Delta\delta$ in detecting helices compared to sheets and coils. This was true for both the AQ and TM protein datasets and with all five random coil tables (Table 6; Fig. 3).

In an attempt to increase predictive accuracy of $\Delta\delta$ for non-helical structure, we performed iterative optimization

of the thresholds for each RCCS table and each nucleus (Table 7; detailed results in the Supplementary Material). Threshold optimization was performed on the AQ dataset by comparing the number of correctly to incorrectly assessed amino acids over a range of thresholds for each nucleus. The total number of ‘correct–incorrect’ was normalized over the total number of predicted amino acids and the highest value was deemed the optimized value. These optimized thresholds are very different, in some cases, from the values proposed by Wishart et al. (1995a) and they vary by nucleus and RCCS dataset. This suggests that the previous thresholds may be too general. As expected, the trend of chemical shift shielding and deshielding relative to secondary structure is the same regardless of the actual threshold value on the nucleus type: H^α and C^β experience shielding in helices and deshielding in sheets, while the opposite trend is true for C^α . cursory examination of Table 6 implies overall higher accuracy of $\Delta\delta$ for predicting secondary structure in TM proteins (~87–88% accuracy) vs. AQ (~62–64% accuracy) proteins. However, the reason for this is actually the predominance of helical structure in the H^α containing portion of the TM dataset (Table 8), rather than an inherently better ability to predict TM protein structure from chemical shifts.

Addition of localized averaging for both the PDB file based dihedral angle test and $\Delta\delta$ analysis with both the optimized and original thresholds increased the overall

Table 8 Number of residues in each dataset categorized by chemical shift availability per residue and classified by secondary structure (helix (H), sheet (S), coil (C)), with secondary structure determined using the consensus structure according to ϕ and ψ dihedral angles over a 5-residue local average around the residue in question

	TM				AQ			
	Total	H	S	C	Total	H	S	C
H^α								
# of files	25				130			
# of Proline	3	3	0	0	418	171	198	49
# of glycines	33	33	0	0	698	242	121	335
# total amino acids	343	335	2	6	10,014	5,176	3,308	1,530
C^α								
# of files	25				99			
# of Proline	19	6	10	3	402	165	189	48
# of glycines	80	37	39	4	677	228	121	328
# total amino acids	606	338	241	27	9,526	4,954	3,084	1,488
C^β								
# of files	25				130			
# of Proline	17	5	9	3	403	166	188	49
# of glycines	0	0	0	0	0	0	0	0
# total amino acids	500	274	203	23	8,777	4,702	2,939	1,136
≥2 of H^α, C^α, C^β								
# of files	19				109			
# of Proline	19	6	10	3	294	104	151	39
# of glycines	80	37	39	4	453	130	94	229
# total amino acids	500	274	203	23	6,127	2,855	2,269	1,003

accuracy of secondary structure prediction. For helix and sheet regions, accuracy generally increased modestly (~2–10%) for both the AQ and TM datasets, but by 20–40% for coils. This implies that structural averaging is important for use of $\Delta\delta$ values, particularly in identification of regions lacking defined secondary structure where individual residues may have characteristics of a secondary structure but where the segment as a whole is a coil. Only subtle differences in accuracy were observed between the original thresholds and our iteratively determined thresholds. This could be attributed to the fact that $\Delta\delta$ values are rather large in comparison to the thresholds for most instances of secondary structuring. Clearly, structural thresholds would be significantly more important for small values of $\Delta\delta$, since these would lie closest to the boundary for prediction of coil versus structured. In these cases, the structural thresholds make a drastic difference. The predictions of sheets in both data sets were relatively poor even when using structural averaging. In comparison to

helical structure, sheets are able to assume a much greater variety of backbone dihedral angles (Lovell et al. 2003), which may be a major factor in the relative inaccuracy in prediction of sheets since this should lead to increased variability in chemical shift perturbation from residue to residue. Furthermore, consideration and preferential weighting of different nuclei, such as those identified by Wang and Jardetzky (2002), may improve differentiation between sheet and coil. The addition of the capability to allow differential use of various chemical shift types to distinguish helix, coil and sheet to CS-CHEMeleon is likely in future iterations, but was not the focus of the present work.

Interestingly, there was no major improvement in accuracy when using RCCS acquired in different conditions when predicting protein secondary structure in the TM dataset versus the AQ dataset. In other words, there is no obvious correlation between the solvent environment used for RCCS determination and protein secondary structure. The RCCS providing the best overall predictive accuracy were those derived in DMSO and those determined in 8 M aqueous urea by Schwarzingler et al. (2000) (Table 6), while the others fell close behind. A study by Mielke and Krishnan (2004), evaluating only non-membrane proteins, came to the same conclusion about the applicability of the chemical shifts of Schwarzingler et al. versus other aqueous RCCS tables.

Although it would be most logical to correlate RCCS environment to the environment of the studied protein (i.e., DMSO or trisolvant with TM proteins and the table of Schwarzingler et al. for AQ proteins), our statistical analysis suggests that this is not actually the case. It is possible that RCCS derived in solvents with much lower ϵ or with complete lack of H-bonding capability, given an appropriately engineered peptide series, would provide significant improvement in secondary structure prediction for membrane proteins or the hydrophobic core region of globular proteins. However, the apparent insensitivity of $\Delta\delta$ -based prediction to fairly dramatic changes in RCCS does not provide a strong incentive to perform these studies. Furthermore, helical regions are already very well predicted, and there may be no major improvement in ability to predict sheet or coil regions, even with a set of chemical shifts derived in such an environment. Based upon our findings, the optimal method for any type of protein being studied in an aqueous or non-aqueous environment is to perform a comparison of the $\Delta\delta$ values derived using both the DMSO-based table and the table of Schwarzingler et al. (2000) in order to determine a consensus for secondary structure prediction. The comparison should also be made with more than one nucleus if possible. This type of comparative analysis is readily possible using CS-CHEMeleon.

Conclusions

We have determined the ^1H and ^{13}C chemical shifts of 21 hexapeptides with sequence Ac-GGXAGG-NH₂, where X is one of either the 20 naturally occurring amino acids or the modified amino acid 4-hydroxyproline, demonstrating significant differences in RCCS between aqueous environments and intermediate ϵ environments. Structural studies of TM proteins that use $\Delta\delta$ values for fast assessment of secondary structure have used RCCS measured in aqueous environments, but the predictive accuracy of $\Delta\delta$ utilization has never been evaluated over a range of solvent environments. In this paper, we provide evidence that, although $\Delta\delta$ values themselves are affected, the solvent in which the RCCS were measured does not significantly affect the prediction of secondary structure. Rather, the type of secondary structure is a major factor in the agreement between $\Delta\delta$ -based prediction and experimental secondary structures. For a thorough assessment, all three major nuclei (H^α , C^α , C^β) should be considered as well as the type of secondary structure being evaluated. Although $\Delta\delta$ use is well suited to an overall estimate of protein structure, the evidence presented herein of a bias towards helices with all RCCS datasets provides a strong incentive to employ alternative restraints during structure calculation and to ensure that $\Delta\delta$ -based restraints are not overrepresented in the energy expression being used for structure calculation. Choice of the $\Delta\delta$ threshold for helix versus coil or sheet also may slightly improve overall secondary structure prediction accuracy. Furthermore, optimal accuracy in secondary structure prediction is probably obtainable by comparison of $\Delta\delta$ obtained with the table presented herein in DMSO and the table determined in aqueous conditions by Schwarzsinger et al. (2000). The web-based software CS-CHEMeleon, introduced herein, provides a rapid and versatile method to allow such a comparative analysis.

Acknowledgments Thanks to Daniel Gaston for helpful discussions; to David Langelaan and Tyler Reddy for critical reading of the manuscript; to Bruce Stewart for technical support; Dr. David Wasman for CD spectropolarimeter access; Dr. Barbara Karten for access to a refrigerated centrifuge; Dr. Carmichael Wallace for access to a lyophilizer; and to Drs. Mike Lumsden, (NMR-3, Dalhousie University), Kathy Robertson (NMR-3), Ray Syvitski (NRC-IMB) and Ian Burton (NRC-IMB) for assistance during configuration of NMR experiments. This work was funded by a Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grant alongside partial stipend support for MLT from a Dalhousie Cancer Research Program grant to JKR. AWB received support as an NSERC Undergraduate Student Research Associate. Critical infrastructure was provided by a Faculty of Medicine Intramural Grant and startup funding from Dalhousie University. Operation of NMR-3 is funded by Dalhousie University and NSERC. The 700 MHz cryoprobe at the NRC-IMB was funded by an Atlantic Canada Opportunities Agency Grant to Dalhousie University.

References

- Abraham RJ, Cavalli L, Pachler KGR (1966) Rotational isomerism.2. A calculation of solvent dependence of conformational equilibria in substituted ethanes and its application to NMR spectra. *Mol Phys* 11:471–481
- Amirjehed AK, Blake MI (1975) Deviation of dielectric constant from ideality for certain binary solvent systems. *J Pharm Sci* 64:1569–1570
- Avbelj F, Kocjan D, Baldwin RL (2004) Protein chemical shifts arising from alpha-helices and beta-sheets depend on solvent exposure. *Proc Natl Acad Sci USA* 101:17394–17397
- Beger RD, Bolton PH (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
- Berjanskii M, Wishart D (2007) The RCI server: rapid and accurate calculation of protein flexibility using chemical shifts. *Nucleic Acids Res* 35:W531–W537
- Berjanskii MV, Neal S, Wishart DS (2006) PREDITOR: a web server for predicting protein torsion angle restraints. *Nucleic Acids Res* 34:W63–W69
- 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
- Blaber M, Zhang XJ, Matthews BW (1993) Structural basis of amino acid alpha-helix propensity. *Science* 260:1637–1640
- Brockman H (1994) Dipole potential of lipid-membranes. *Chem Phys Lipids* 73:57–79
- Burger R, Bigler P (1998) DEPTQ: distortionless enhancement by polarization transfer including the detection of quaternary nuclei. *J Magn Reson* 135:529–534
- Case DA (1998) The use of chemical shifts and their anisotropies in biomolecular structure determination. *Curr Opin Struct Biol* 8:624–630
- Case DA (2000) Interpretation of chemical shifts and coupling constants in macromolecules. *Curr Opin Struct Biol* 10:197–203
- Cavalli A, Salvatella X, Dobson CM, Vendruscolo M (2007) Protein structure determination from NMR chemical shifts. *Proc Natl Acad Sci USA* 104:9615–9620
- Chellgren BW, Creamer TP (2004) Short sequences of non-proline residues can adopt the polyproline II helical conformation. *Biochemistry* 43:5864–5869
- Cheung M, Maguire ML, Stevens TJ, Broadhurst RW (2010) DANGLE: a Bayesian inferential method for predicting protein backbone dihedral angles and secondary structure. *J Magn Reson* 202:223–233
- Cornilescu G, Delaglio F, Bax A (1999) Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J Biomol NMR* 13:289–302
- Dalgarno DC, Levine BA, Williams RJP (1983) Structural information from NMR secondary chemical-shifts of peptide alpha-C-H protons in proteins. *Biosci Rep* 3:443–452
- Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A (1995) NMRPipe—a multidimensional spectral processing system based on unix pipes. *J Biomol NMR* 6:277–293
- Evans JNS (1995) Biomolecular NMR spectroscopy. Oxford University Press, New York
- Farrow NA, Zhang OW, Formankay JD, Kay LE (1994) A heteronuclear correlation experiment for simultaneous determination of N-15 longitudinal decay and chemical-exchange rates of systems in slow equilibrium. *J Biomol NMR* 4:727–734
- Garciamoreno B, Dwyer JJ, Gittis AG, Lattman EE, Spencer DS, Stites WE (1997) Experimental measurement of the effective

- dielectric in the hydrophobic core of a protein. *Biophys Chem* 64:211–224
- Goddard TD, Kneller DG (2008) *Sparky 3*. University of California, San Francisco
- Greenfield NJ (2004) Analysis of circular dichroism data. *Method Enzymol* 383:282–317
- Guy CA, Fields GB (1997) Trifluoroacetic acid cleavage and deprotection of resin-bound peptides following synthesis by Fmoc chemistry. *Method Enzymol* 289:67–83
- Hass MAS, Jensen MR, Led JJ (2008) Probing electric fields in proteins in solution by NMR spectroscopy. *Proteins* 72:333–343
- Hoffman R (2003) Variations on the chemical shift of TMS. *J Magn Reson* 163:325–331
- Hoffman R (2006) Standardization of chemical shifts of TMS and solvent signals in NMR solvents. *Magn Reson Chem* 44: 606–616
- Hutchinson EG, Thornton JM (1996) PROMOTIF—a program to identify and analyze structural motifs in proteins. *Protein Sci* 5:212–220
- Iwatake M, Asakura T, Williamson MP (1999) C-alpha and C-beta carbon-13 chemical shifts in proteins from an empirical database. *J Biomol NMR* 13:199–211
- Jiao D, Barfield M, Hruby V (1993) Ab initio IGLO study of the phi and chi-angle dependence of the carbon-13 chemical shifts in the model peptide N-acetyl-N'-methylglycinamide. *J Am Chem Soc* 115:10883–10887
- Kim H, Gao J, Burgess DJ (2009) Evaluation of solvent effects on protonation using NMR spectroscopy: implication in salt formation. *Int J Pharm* 377:105–111
- Koehorst RB, Spruijt RB, Hemminga MA (2008) Site-directed fluorescence labeling of a membrane protein with BADAN: probing protein topology and local environment. *Biophys J* 94:3945–3955
- Kuszewski J, Gronenborn AM, Clore GM (1996) A potential involving multiple proton chemical-shift restraints for nonsterespecifically assigned methyl and methylene protons. *J Magn Reson Ser B* 112:79–81
- Langelaa D, Bebbington ME, Reddy T, Rainey JK (2009) Structural insight into G-protein coupled receptor binding by apelin. *Biochemistry* 48:537–548
- Le HB, Oldfield E (1994) Correlation between N-15 NMR chemical-shifts in proteins and secondary structure. *J Biomol NMR* 4: 341–348
- Li SC, Deber CM (1994) A measure of helical propensity for amino acids in membrane environments. *Nat Struct Biol* 1:368–373
- Lovell SC, Davis IW, Adrendall WB, De Bakker PIW, Word JM, Prisant MG, Richardson JS, Richardson DC (2003) Structure validation by C alpha geometry: phi, psi and C beta deviation. *Proteins* 50:437–450
- Lukin JA, Gove AP, Talukdar SN, Ho C (1997) Automated probabilistic method for assigning backbone resonances of (13C, 15N)-labeled proteins. *J Biomol NMR* 9:151–166
- Merutka G, Dyson HJ, Wright PE (1995) 'Random coil' 1H chemical shifts obtained as a function of temperature and trifluoroethanol concentration for the peptide series GGXGG. *J Biomol NMR* 5:14–24
- Mielke SP, Krishnan VV (2004) An evaluation of chemical shift index-based secondary structure determination in proteins: influence of random coil chemical shifts. *J Biomol NMR* 30:143–153
- Osapay K, Case DA (1991) A new analysis of proton chemical shifts in proteins. *J Am Chem Soc* 113:9436–9444
- Plaxco K, Morton C, Grimshaw S, Jones J, Pitkeathly M, Campbell I, Dobson C (1997) The effects of guanidine hydrochloride on the 'random coil' conformations and NMR chemical shifts of the peptide series GGXGG. *J Biomol NMR* 10:221–230
- Quirt A, Lyerla J, Peat I, Cohen J, Reynolds W, Freedman M (1974) C-13 Nuclear magnetic-resonance titration shifts in amino-acids. *J Am Chem Soc* 96:570–571
- Rainey J, Fliegel L, Sykes BD (2006) Strategies for dealing with conformational sampling in structural calculations of flexible or kinked transmembrane peptides. *Biochem Cell Biol* 84:918–929
- Rath A, Davidson AR, Deber CM (2005) The structure of "unstructured" regions in peptides and proteins: role of the polyproline II helix in protein folding and recognition. *Biopolymers* 80: 179–185
- Schwarzinger S, Kroon GJA, Foss TR, Wright PE, Dyson HJ (2000) Random coil chemical shifts in acidic 8 M urea: implementation of random coil shift data in NMRView. *J Biomol NMR* 18: 43–48
- Schwarzinger S, Kroon GJA, Foss TR, Chung J, Wright PE, Dyson HJ (2001) Sequence-dependent correction of random coil NMR chemical shifts. *J Am Chem Soc* 123:2970–2978
- Shen Y, Lange O, Delaglio F, Rossi P, Aramini JM, Liu G, Eletsky A, Wu Y, Singarapu KK, Lemak A, Ignatchenko A, Arrowsmith CH, Szyperski T, Montelione GT, Baker D, Bax A (2008) Consistent blind protein structure generation from NMR chemical shift data. *Proc Natl Acad Sci USA* 105:4685–4690
- Shen Y, Delaglio F, Cornilescu G, Bax A (2009a) TALOS plus: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J Biomol NMR* 44:213–223
- Shen Y, Vernon R, Baker D, Bax A (2009b) De novo protein structure generation from incomplete chemical shift assignments. *J Biomol NMR* 43:63–78
- Shenderovich IG, Burtsev AP, Denisov GS, Golubev NS, Limbach HH (2001) Influence of the temperature-dependent dielectric constant on the H/D isotope effects on the NMR chemical shifts and the hydrogen bond geometry of the collidine-HF complex in CDF3/CDCIF2 solution. *Magn Reson Chem* 39:S91–S99
- Slepkov ER, Rainey JK, Li XJ, Liu YS, Cheng FJ, Lindhout DA, Sykes BD, Fliegel L (2005) Structural and functional characterization of transmembrane segment IV of the NHE1 isoform of the Na+/H+ exchanger. *J Biol Chem* 280:17863–17872
- Sondergaard CR, Mcintosh LP, Pollastri G, Nielsen JE (2008) Determination of electrostatic interaction energies and protonation state populations in enzyme active sites. *J Mol Biol* 376:269–287
- Spera S, Bax A (1991) Empirical correlation between protein backbone conformation and C-alpha and C-beta C-13 nuclear magnetic resonance chemical shifts. *J Am Chem Soc* 113:5490–5492
- Szilagy L, Jardetzky O (1989) Alpha-protons chemical-shifts and secondary structure in proteins. *J Magn Reson* 83:441–449
- Thanabal V, Omecinsky DO, Reily MD, Cody WL (1994) The 13C chemical shifts of amino acids in aqueous solution containing organic solvents: application to the secondary structure characterization of peptides in aqueous trifluoroethanol solution. *J Biomol NMR* 4:47–59
- Tonan K, Ikawa S (2003) Effect of solvent on an NMR chemical shift difference between glycylic geminal alpha-protons as a probe of beta-turn formation of short peptides. *Spectrochim Acta A Mol Biomol Spectrosc* 59:111–120
- Tusnady GE, Dosztanyi Z, Simon I (2005) PDB_TM: selection and membrane localization of transmembrane proteins in the protein data bank. *Nucleic Acids Res* 33:D275–D278
- Ulrich EL, Akutsu H, Doreleijers JF, Harano Y, Ioannidis YE, Lin J, Livny M, Mading S, Maziuk D, Miller Z, Nakatani E, Schulte CF, Tolmie DE, Wenger RK, Yao H, Markley JL (2008) BioMagResBank. *Nucleic Acids Res* 36:D402–D408
- Wang Y, Jardetzky O (2002) Probability-based protein secondary structure identification using combined NMR chemical-shift data. *Protein Sci* 11:852–861

- Williamson MP (1990) Secondary-structure dependent chemical-shifts in proteins. *Biopolymers* 29:1428–1431
- Williamson MP, Asakura T, Nakamura E, Demura M (1992) A method for the calculation of protein alpha-CH chemical-shifts. *J Biomol NMR* 2:83–98
- Wishart DS, Sykes BD (1994) The C-13 chemical-shift index—A simple method for the identification of protein secondary structure using C-13 chemical-shift data. *J Biomol NMR* 4:171–180
- Wishart DS, Sykes BD, Richards FM (1992) The chemical shift index—a fast and simple method for the assignment of protein secondary structure through NMR-spectroscopy. *Biochemistry* 31:1647–1651
- Wishart DS, Bigam CG, Holm A, Hodges RS, Sykes BD (1995a) H-1, C-13 and N-15 random coil NMR chemical shifts of the common amino acids.1. Investigations of nearest-neighbor effects. *J Biomol NMR* 5:67–81
- Wishart DS, Bigam CG, Yao J, Abildgaard F, Dyson HJ, Oldfield E, Markley JL, Sykes BD (1995b) H-1, C-13 and N-15 chemical shifts referencing in biomolecular NMR. *J Biomol NMR* 6:135–140
- Wishart DS, Arndt D, Berjanskii M, Tang P, Zhou J, Lin G (2008) CS23D: a web server for rapid protein structure generation using NMR chemical shifts and sequence data. *Nucleic Acids Res* 36:W496–W502
- Wüthrich K (1986) *NMR of proteins and nucleic acids*. Wiley, United States, p 17
- Xu XP, Case DA (2001) Automated prediction of N-15, C-13(alpha), C-13(beta) and C-13 'chemical shifts in proteins using a density functional database. *J Biomol NMR* 21:321–333
- Zuiderweg ERP, Nettlesheim DG, Mollison KW, Carter GW (1989) Tertiary structure of human-complement component C5A in solution from nuclear magnetic-resonance data. *Biochemistry* 28:172–185