

Improved chemical shift prediction by Rosetta conformational sampling

Ye Tian · Stanley J. Opella · Francesca M. Marassi

Received: 12 July 2012 / Accepted: 16 September 2012 / Published online: 25 September 2012
© Springer Science+Business Media Dordrecht 2012

Abstract Chemical shift frequencies represent a time-average of all the conformational states populated by a protein. Thus, chemical shift prediction programs based on sequence and database analysis yield higher accuracy for rigid rather than flexible protein segments. Here we show that the prediction accuracy can be significantly improved by averaging over an ensemble of structures, predicted solely from amino acid sequence with the Rosetta program. This approach to chemical shift and structure prediction has the potential to be useful for guiding resonance assignments, especially in solid-state NMR structural studies of membrane proteins in proteoliposomes.

Keywords Chemical shift · Prediction · Rosetta · Ensemble averaging · Structure · Membrane protein · Solid-state NMR · Conformational ensemble

NMR methods for protein structure determination increasingly rely on measurements of dipolar coupling, chemical shift anisotropy and isotropic chemical shift to provide molecular orientation and backbone dihedral angle (ϕ , ψ) restraints (Delaglio et al. 2000; Tian et al. 2001; Marassi and Opella 2003; Asbury et al. 2006; Cavalli et al. 2007; Shen et al. 2008; Wishart et al. 2008). This approach

relieves the time-consuming burden of measuring multiple, long-range distances from side chains, and facilitates solution NMR and solid-state NMR structure determination by molecular fragment replacement for soluble proteins in water (Shen et al. 2008; Raman et al. 2010), membrane proteins in micelles (Berardi et al. 2011) and membrane proteins in lipid bilayers (Das et al. 2012; Marassi and Opella 2003; Sharma et al. 2010; Park et al. 2012).

While dipolar couplings and chemical shift anisotropies have the distinct advantage of directly reflecting molecular orientation relative to the direction of the magnetic field or the lipid bilayer membrane, isotropic chemical shifts are also important sources of structural information (Clayden and Williams 1982; Dalgarno et al. 1983; Saito 1986; Gross and Kalbitzer 1988; Szilagy and Jardetzky 1989; Pastore and Saudek 1990; Osapay and Case 1991; Spera and Bax 1991; Wishart et al. 1991; Iwadate et al. 1999). Several approaches have been developed both for deriving backbone dihedral angles from chemical shifts (Luginbuhl et al. 1995; Cornilescu et al. 1999) and for generating protein structures using chemical shifts as the sole source of experimental restraints, including CHESHIRE (Cavalli et al. 2007), CS23D (Wishart et al. 2008) and CS-Rosetta (Shen et al. 2008; Shen et al. 2009).

The reverse process, where isotropic chemical shifts are predicted from a structural model is also possible. Chemical shifts can be calculated ab initio from molecular structures (de Dios et al. 1993; Xu and Case 2001; Vila et al. 2008; Vila et al. 2009) and methods based on neural networks also make good chemical shift predictions (Meller 2003). However, methods based on sequence homology and database analysis, such as ShiftX (Neal et al. 2003), ShiftX2 (Han et al. 2011), SPARTA (Shen and Bax 2007), SPARTA+ (Shen and Bax 2010), CamShift (Kohlhoff et al. 2009), and shAIC (Nielsen et al. 2012),

Y. Tian · F. M. Marassi (✉)
Sanford Burnham Medical Research Institute, 10901 North
Torrey Pines Road, La Jolla, CA 92037, USA
e-mail: fmarassi@sbmri.org

Y. Tian · S. J. Opella
Department of Chemistry and Biochemistry, University
of California San Diego, 9500 Gilman Drive, La Jolla,
CA 92093-0307, USA

have been shown to yield better agreement between experimentally observed and predicted values.

As proteins are intrinsically dynamic, the experimentally measured isotropic chemical shifts represent both an ensemble and time average over all explored conformational states (Mittag and Forman-Kay 2007). Thus, significant improvements in chemical shift prediction have been obtained either by averaging the predicted chemical shifts over extended ($\sim 1 \mu\text{s}$) molecular dynamics trajectories (Li and Bruschweiler 2009), or by averaging over molecular dynamics trajectories obtained by enhanced conformational space sampling (Markwick et al. 2010).

Here we show that averaging over an ensemble of protein structural models, predicted de novo from their amino acid sequences, can also significantly improve the root-mean-square difference (RMSD) between predicted and experimental chemical shift values. We demonstrate this for two proteins from the bacterial mercury detoxification *Mer* operon: the soluble periplasmic protein MerP, whose structure has been determined by solution NMR in aqueous buffer (Steele and Opella 1997), and the integral membrane protein MerF, whose structure has been determined by solid-state NMR in proteoliposomes (Das et al. 2012), by solid-state NMR in magnetically aligned phospholipid bilayers (De Angelis et al. 2006), and by solution NMR in detergent micelles (Howell et al. 2005).

The Rosetta program is very successful for predicting atomic level three-dimensional structures of proteins from their amino acid sequences (Das and Baker 2008). Starting with the sequences of MerP (all 72 residues) and MerFt (residues 13–72 in the helix-loop-helix integral membrane core of 81-residue MerF), we performed blind structure predictions using Rosetta, by excluding the PDB structural coordinates of the two proteins as well as their homologues from the structure prediction database. For each protein, 10,000 coarse-grained structural models were generated and then refined by all-atom relaxation, performed with either the implicit aqueous environment protocol (for MerP) or the implicit membrane protocol (for MerFt) available in Rosetta 3.2 (Das and Baker 2008; Yarov-Yarovoy et al. 2006). The refined all-atom structures were clustered according to their overall energy and their backbone CA atom RMSD to the lowest energy structure with a cutoff of 5 Å. For each protein, the most populated cluster of molecular ensembles encompassed more than 40 % of the entire sampling space and contained 5,928 structures for MerP and 4,273 structures for MerFt.

For each of these structures, isotropic chemical shifts were predicted using ShiftX2 (v1.03) (Han et al. 2011) and SPARTA+ (Shen and Bax 2010), and then averaged over an increasing number of conformations, ranked from lowest to highest all-atom energy, in selected Rosetta ensembles. The results were compared to the experimental

chemical shift values measured for MerP in aqueous buffer (Steele and Opella 1997) or the values measured with MAS solid-state NMR experiments on MerFt in proteoliposomes (Das et al. 2012).

The protein structures predicted with Rosetta agree very well with those determined by NMR, with backbone RMSDs of 1.4 Å for MerP and 6.0 Å for MerFt (Fig. 1). For MerFt, the higher RMSD reflects primarily conformational differences in the termini of the two structures. However, this does not seem to adversely affect the accuracy of chemical shift prediction. More substantial motions in the terminal regions could help mask any conformational differences through averaging, resulting in both more accurate chemical shift prediction and its improvement upon ensemble averaging (see below).

The chemical shifts measured for backbone ^{15}N and $^{13}\text{C}\alpha$ sites of MerFt in proteoliposomes by solid-state NMR agree remarkably well with those measured from the same protein in micelles by solution NMR (Fig. 2), indicating that the overall transmembrane helical structure of the protein is preserved in the two different environments. It will be interesting to examine whether good correlations can also be observed for lipid-water interfacial helices of membrane proteins where the environment is highly heterogeneous.

Significant correlation has been found to exist between the accuracy of a Rosetta structural model and its agreement with experimental chemical shifts (Shen et al. 2008). Furthermore, the model quality required for accurate chemical shift prediction will likely depend on the extent of local dynamics and on the specific nature of the protein's structure. For example, mixed α - β structures, such as MerP, are more challenging to model than α -helical structures, such as MerFt (Bonneau et al. 2001). Nevertheless, for both MerP and MerFt, a significant improvement in chemical shift prediction accuracy was observed when the predicted frequencies were averaged over an ensemble of conformations, rather than taken from a single lowest energy structural model (Fig. 1d, h; Table 1).

The chemical shifts predicted from the single lowest energy Rosetta structures correlate well with the experimental data, with cumulative RMSDs of 6.2 for MerP (N, HN, CA, CB) and 7.7 for MerFt (N, CA, CO). However, ensemble-averaging of the chemical shifts yielded lower cumulative RMSDs for both proteins, 5.7 for MerP and 6.8 for MerFt, representative of a $\sim 12\%$ improvement over single structure prediction. ShiftX2 (Table 1; Figs. 1, 3, 4) and SPARTA+ (data not shown) yielded similar results, although somewhat better prediction accuracy was obtained with ShiftX2.

Correlation with the experimental data improves as the predicted chemical shifts are averaged over an increasing number of ensemble structures (Fig. 3). The improvement

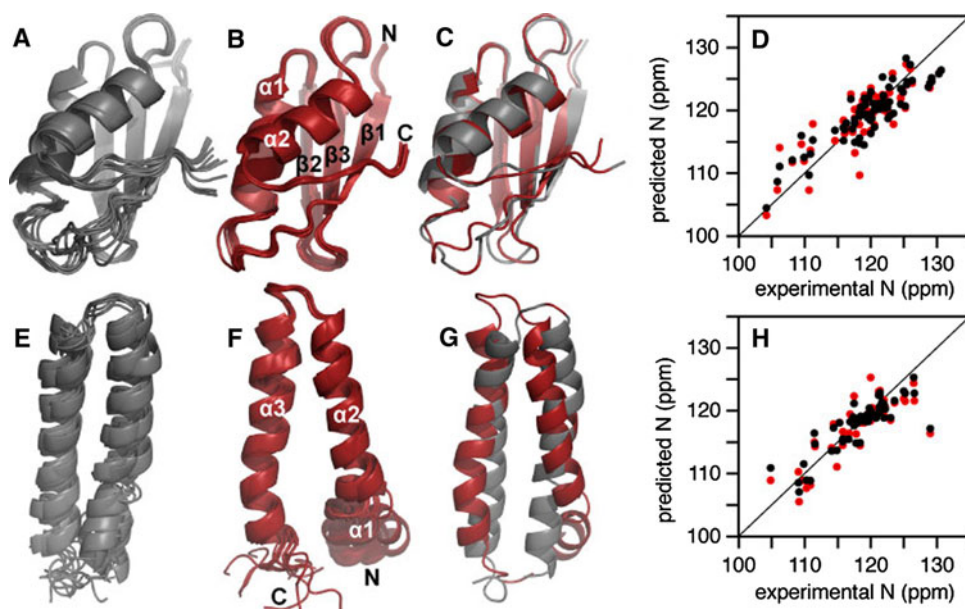


Fig. 1 Conformational ensemble averaging improves chemical shift prediction for the soluble periplasmic protein MerP (a–d) and the integral membrane protein MerFt (e–h). **a, e** Overlay of 10 lowest energy Rosetta structures of MerP and MerFt. **b, f** Overlay of 10 lowest energy structures determined by solution NMR for MerP (PDB: 1AFI) and by solid-state NMR for MerFt (PDB: 2LJ2). **c, g** Overlay of lowest energy structures from Rosetta (gray) and NMR

(red). **d, h** Correlation between experimental and predicted ^{15}N chemical shifts obtained for the single lowest energy Rosetta structure (red), or by averaging over an ensemble of 20 lowest energy Rosetta structures (black). Experimental chemical shifts and NMR coordinates were taken from the published data for MerP in aqueous solution (Steele and Opella 1997) and MerFt in proteoliposomes (Das et al. 2012)

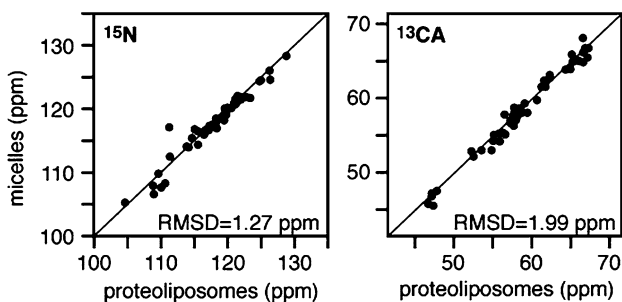


Fig. 2 Correlation between solution and solid-state NMR chemical shifts of MerFt. The ^{15}N and $^{13}\text{C}\alpha$ chemical shifts were measured for MerFt in micelles by solution NMR (Howell et al. 2005) and in proteoliposomes by solid-state NMR (Das et al. 2012). The RMSDs are 1.27 ppm for ^{15}N and 1.99 ppm for $^{13}\text{C}\alpha$

levels off when >10 structural models are included in the averaging process, reflecting the onset of contributions from higher energy conformers and indicating that a Rosetta ensemble of 10 lowest energy structures is a better representation of the native structure than the single lowest energy model.

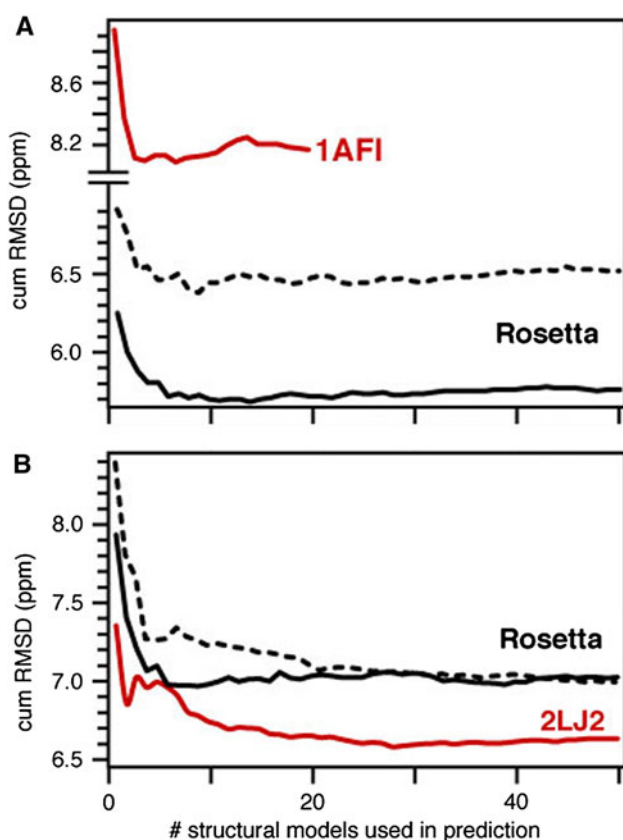
Notably, within a given conformational ensemble, the accuracy of chemical shift prediction is always better when the data are averaged over multiple conformations rather than taken individually, although averaging over the lowest energy structures (Fig. 3, black solid line) yields the most accurate prediction. For example, averaging the predicted

chemical shifts over an ensemble of 100 highest Rosetta energy structures ranked from lowest to highest energy (Fig. 3, black dashed line) also yields an improvement in prediction accuracy. For MerFt, the improvement levels off at the same value as for prediction from the lowest energy ensemble, albeit after including at least twice as many conformations (Fig. 3b). For MerP, ensemble averaging over the higher energy structures improves the prediction but the accuracy never reaches the level obtained for the low energy ensembles (Fig. 3a). This difference between MerP and MerFt may reflect differences in protein dynamics, as well as the distinctly different conformations and environments of the two proteins (compact, soluble α - β structure for MerP; two transmembrane α -helices for MerFt). Since MerP is water-soluble and MerF is an integral membrane protein, their respective models were predicted using either the all atom implicit solvent potential (MerP) or the all atom implicit membrane potential (MerFt) of Rosetta.

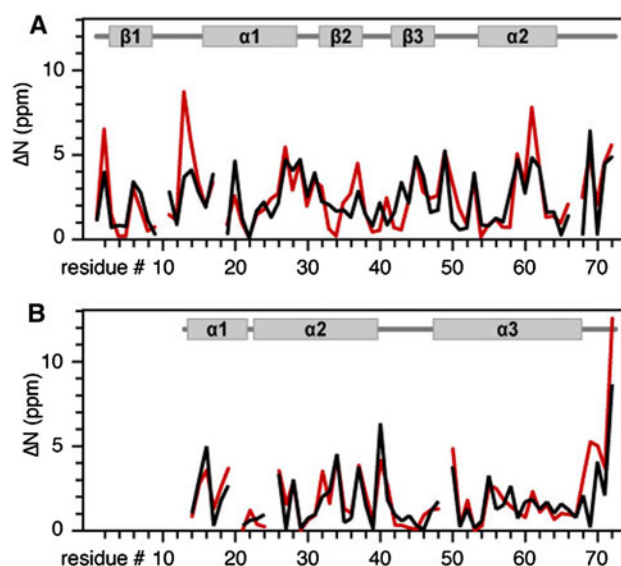
The most notable improvements in prediction accuracy from ensemble averaging were obtained for the ^{15}N nuclei (Table 1). Furthermore, the degree of improvement correlates with protein dynamics (Fig. 4). Similar results were reported in a previous study where the predicted chemical shifts were averaged over molecular dynamics trajectories (Markwick et al. 2010). This is consistent with the observation that chemical shift prediction programs based on

Table 1 RMSDs between experimental and predicted shifts obtained for a single structure ($n = 1$) or by averaging over an ensemble of n structures generated by Rosetta prediction or determined by NMR spectroscopy

Structure	RMSD (ppm)						
	n	HN	N	CA	CB	CO	Cumulative
MerP (Rosetta)	1	0.63	3.10	1.39	1.11		6.2
	20	0.61	2.81	1.23	1.04		5.7
	100	0.61	2.84	1.24	1.05		5.7
MerP (1AFI)	1	0.71	4.93	1.61	1.65		8.9
	20	0.69	4.43	1.49	1.53		8.1
MerFt (Rosetta)	1		2.77	2.31		2.57	7.7
	20		2.53	1.74		2.50	6.8
	100		2.49	1.80		2.49	6.8
MerFt (2LJ2)	1		2.90	1.66		2.54	7.1
	20		2.42	1.59		2.38	6.4
	100		2.40	1.61		2.38	6.4

**Fig. 3** Improvement in cumulative RMSD between experimental and predicted chemical shifts due to ensemble averaging for MerP (a) and MerFt (b). Predicted chemical shifts were averaged over increasing numbers of ensemble structures ranked from lowest to highest energy. (Red) Ensembles of lowest energy NMR structures. (Black solid line) Ensembles of lowest energy Rosetta structures. (Black dashed line) Ensembles of 100 highest energy Rosetta structures

homology and database analysis yield higher prediction accuracy for rigid rather than flexible protein segments, and explains why they can be improved by ensemble averaging

**Fig. 4** Residue-specific improvement in ^{15}N chemical shift prediction accuracy obtained by ensemble averaging for MerP (a) and MerFt (b). The improvement value (ΔN) is the difference between experimentally observed chemical shifts and chemical shifts predicted from single lowest energy Rosetta structure (red), or chemical shifts predicted and averaged over 20 lowest energy Rosetta structures (black). Protein secondary structures are shown above the graphs

(Shen and Bax 2010). For MerP (Fig. 4a) the largest improvements in prediction accuracy coincide with residues in the $\beta 1$ - $\alpha 1$ metal-binding loop, which is more flexible than the rest of the protein (Steele and Opella 1997). For MerFt (Fig. 4b), prediction improvements coincide primarily with the more flexible C-terminus of the protein.

The effects of side-chain conformation on backbone chemical shifts are well recognized (de Dios et al. 1993; Wang and Jardetzky 2004; Vila et al. 2007; Villegas et al. 2007; London et al. 2008; Mulder 2009). For example,

excluding χ^1 torsion angle input information from SPARTA+ chemical shift prediction has been shown to decrease the agreement with experimental data by 5 % for ^{15}N and by 1–2 % for other nuclei (Shen and Bax 2010). In the case of MerFt, an additional 5 % improvement in prediction accuracy is observed when the chemical shifts are predicted for and averaged over the experimentally determined solid-state NMR conformers (Fig. 3b, red). This likely coincides with refinement of side chain conformations by the NMR-restrained simulated annealing performed during structure refinement (Das et al. 2012). In the case of MerP, the accuracy of chemical shifts predicted from the NMR structure is poorer than that obtained from Rosetta (Fig. 3a, red). This reflects local differences in backbone dihedral angles between the NMR and Rosetta structures of MerP, found primarily in the $\beta 1$ – $\alpha 1$ and $\alpha 1$ – $\beta 2$ loops (Fig. 5a, b) and may be due to a localized distortion in the NMR structure. Conversely for MerFt, the local dihedral angles of the Rosetta and NMR structures are very similar (Fig. 5c, d) even though the overall backbone RMSD is poorer. The results in Figs. 3 and 5 indicate that local dihedral angles can have a substantial impact on the accuracy of chemical shift prediction even in cases, such as MerP, where the overall fold is accurate.

In summary, we show that the accuracy of chemical shift prediction can be significantly improved by averaging over an ensemble of conformations predicted *de novo* solely from the amino acid sequence. Ensemble averaging

improves the prediction accuracy both for solution NMR data from a globular protein in water and for solid-state NMR data from an integral membrane protein in proteoliposomes. The resulting chemical shift prediction accuracy (cumulative RMSD ~ 6 ppm) is comparable to the accuracy obtained by ensemble averaging over MD trajectories of crystal structures (Markwick et al. 2010). It is not clear whether this accuracy is sufficient to assign the resonances of an entire protein, nonetheless, it definitely has the potential to resolve ambiguities and provide validation in the early stages of experimental structural determination by NMR.

Estimates of the protein size and type compatible with *de novo* Rosetta structure prediction vary with protein secondary structure and the exact methods used in prediction (Das and Baker 2008). Small α -helical proteins (up to ~ 120 residues) are typically well within the size and complexity limits of Rosetta, while mixed α – β structures are more challenging to model (Bonneau et al. 2001). The incorporation of chemical shifts in Rosetta structure prediction has been shown to dramatically improved correct structure convergence for proteins up to 100 residues (Shen et al. 2008) and the addition of RDCs further improves convergence for proteins up to 200 residues (Raman et al. 2010). Rosetta structure prediction works well for many proteins up to 100 amino acids using sequence information alone (Bradley et al. 2005) and for significantly larger proteins when assisted by a template (Das et al. 2007). The estimates

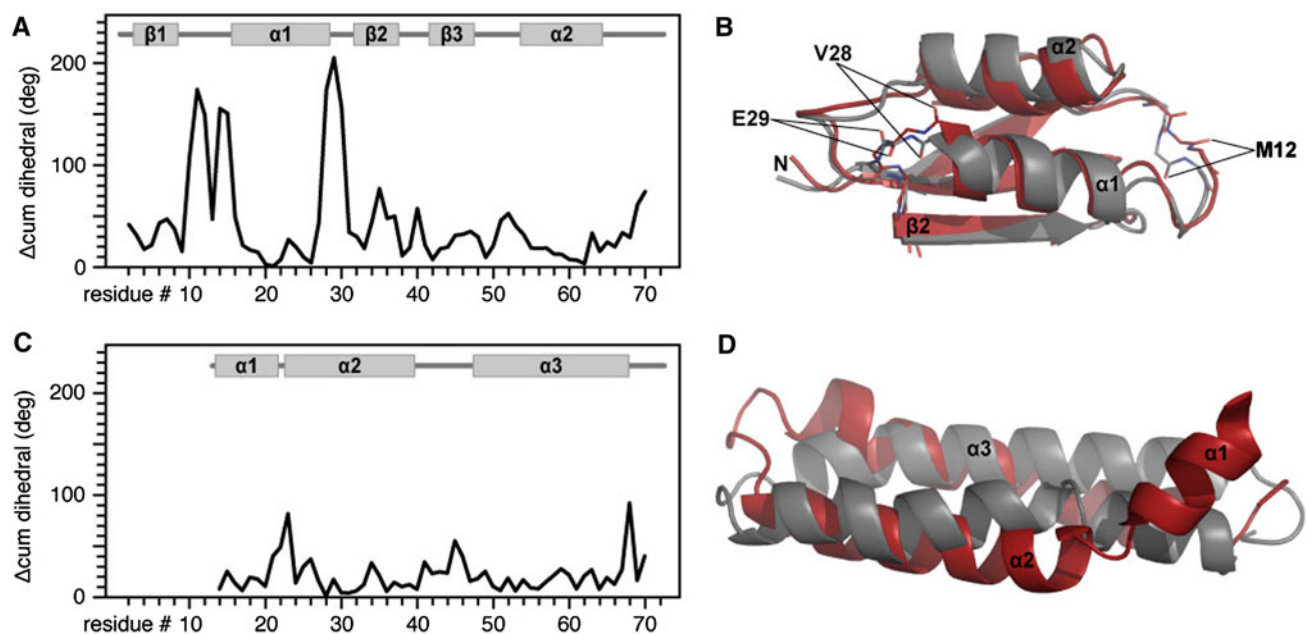


Fig. 5 Agreement between backbone dihedral angles predicted from Rosetta or determined experimentally for MerP (**a, b**) and MerFt (**c, d**). **a, c** Cumulative difference between the backbone dihedral angles (ϕ , ψ) determined by NMR and predicted from Rosetta

($\Delta_{\text{cum}} = \Delta\phi + \Delta\psi$). **b, d** Overlay of NMR (gray) and Rosetta (red) structures. Dihedral angles for MerP residues M12, V28 and E29 differ significantly between the NMR and Rosetta structures

may be even better for membrane proteins where the lipid bilayer membrane that is modeled implicitly in Rosetta (Yarov-Yarovoy et al. 2006) provides additional conformational and spatial restraints. Furthermore, recent methods for structure prediction based on evolutionary coupling can be used to provide very good models of both soluble and integral membrane proteins (Marks et al. 2011; Hopf et al. 2012).

In cases where good structural models can be predicted from amino acid sequence alone, or with assistance of a template derived from sequence homology, or by evolutionary coupling methods, the ensemble-averaging approach described here could provide useful information at a very early stage of a protein structure determination project. In particular, it could serve as a useful guide and validation tool for resonance assignment during de novo solid-state NMR structure determination of membrane proteins, where the process is complicated by broader or overlapped lines and, typically, the absence of ^1H signals, and especially when combined with recently developed automated assignments methods (Moseley et al. 2010; Tycko and Hu 2010).

Acknowledgments This research was supported by grants from the National Institutes of Health (R21GM094727; P01AI074805). It utilized the Biotechnology Research Center for NMR Molecular Imaging of Proteins at UCSD (P41EB002031).

References

- Asbury T, Quine JR, Achuthan S, Hu J, Chapman MS, Cross TA, Bertram R (2006) PIPATH: an optimized algorithm for generating alpha-helical structures from PISEMA data. *J Magn Reson* 183(1):87–95
- Berardi MJ, Shih WM, Harrison SC, Chou JJ (2011) Mitochondrial uncoupling protein 2 structure determined by NMR molecular fragment searching. *Nature* 476(7358):109–113. doi:10.1038/nature10257
- Bonneau R, Tsai J, Ruczinski I, Chivian D, Rohl C, Strauss CE, Baker D (2001) Rosetta in CASP4: progress in ab initio protein structure prediction. *Proteins Suppl* 5:119–126. doi:10.1002/prot.1170
- Bradley P, Misura KM, Baker D (2005) Toward high-resolution de novo structure prediction for small proteins. *Science* 309(5742):1868–1871. doi:10.1126/science.1113801
- Cavalli A, Salvatella X, Dobson CM, Vendruscolo M (2007) Protein structure determination from NMR chemical shifts. *Proc Natl Acad Sci USA* 104(23):9615–9620. doi:10.1073/pnas.0610313104
- Clayden NJ, Williams RJP (1982) Peptide group shifts. *J Magn Reson* 49(3):383–396. doi:10.1016/0022-2364(82)90252-9
- 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(3):289–302
- Dalgarno DC, Levine BA, Williams RJ (1983) Structural information from NMR secondary chemical shifts of peptide alpha C-H protons in proteins. *Biosci Rep* 3(5):443–452
- Das R, Baker D (2008) Macromolecular modeling with Rosetta. *Annu Rev Biochem* 77:363–382. doi:10.1146/annurev.biochem.77.062906.171838
- Das R, Qian B, Raman S, Vernon R, Thompson J, Bradley P, Khare S, Tyka MD, Bhat D, Chivian D, Kim DE, Sheffler WH, Malmstrom L, Wollacott AM, Wang C, Andre I, Baker D (2007) Structure prediction for CASP7 targets using extensive all-atom refinement with Rosetta at home. *Proteins* 69(Suppl 8):118–128. doi:10.1002/prot.21636
- Das BB, Nothnagel HJ, Lu GJ, Son WS, Tian Y, Marassi FM, Opella SJ (2012) Structure determination of a membrane protein in proteoliposomes. *J Am Chem Soc* 134(4):2047–2056. doi:10.1021/ja209464f
- De Angelis AA, Howell SC, Nevzorov AA, Opella SJ (2006) Structure determination of a membrane protein with two transmembrane helices in aligned phospholipid bicelles by solid-state NMR spectroscopy. *J Am Chem Soc* 128(37):12256–12267. doi:10.1021/ja063640w
- de Dios AC, Pearson JG, Oldfield E (1993) Secondary and tertiary structural effects on protein NMR chemical shifts: an ab initio approach. *Science* 260(5113):1491–1496
- Delaglio F, Kontaxis G, Bax A (2000) Protein structure determination using molecular fragment replacement and NMR dipolar couplings. *J Am Chem Soc* 122:2142–2143
- Gross K-H, Kalbitzer HR (1988) Distribution of chemical shifts in ^1H nuclear magnetic resonance spectra of proteins. *J Magn Reson* 76(1):87–99. doi:10.1016/0022-2364(88)90203-x
- Han B, Liu Y, Ginzinger SW, Wishart DS (2011) SHIFTX2: significantly improved protein chemical shift prediction. *J Biomol NMR* 50(1):43–57. doi:10.1007/s10858-011-9478-4
- Hopf TA, Colwell LJ, Sheridan R, Rost B, Sander C, Marks DS (2012) Three-dimensional structures of membrane proteins from genomic sequencing. *Cell* 149(7):1607–1621. doi:10.1016/j.cell.2012.04.012
- Howell SC, Mesleh MF, Opella SJ (2005) NMR structure determination of a membrane protein with two transmembrane helices in micelles: MerF of the bacterial mercury detoxification system. *Biochemistry* 44(13):5196–5206. doi:10.1021/bi048095v
- Iwadata 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(3):199–211
- Kohlhoff KJ, Robustelli P, Cavalli A, Salvatella X, Vendruscolo M (2009) Fast and accurate predictions of protein NMR chemical shifts from interatomic distances. *J Am Chem Soc* 131(39):13894–13895. doi:10.1021/ja903772t
- Li D-W, Bruschweiler R (2009) Certification of molecular dynamics trajectories with NMR chemical shifts. *J Phys Chem Lett* 1(1):246–248. doi:10.1021/jz9001345
- London RE, Wingad BD, Mueller GA (2008) Dependence of amino acid side chain ^{13}C shifts on dihedral angle: application to conformational analysis. *J Am Chem Soc* 130(33):11097–11105. doi:10.1021/ja802729t
- Luginbuhl P, Szyperski T, Wuthrich K (1995) Statistical basis for the use of ^{13}C chemical shifts in protein structure determination. *J Magn Reson B* 109(2):229–233. doi:10.1006/jmrb.1995.0016
- Marassi FM, Opella SJ (2003) Simultaneous assignment and structure determination of a membrane protein from NMR orientational restraints. *Protein Sci* 12(3):403–411. doi:10.1110/ps.0211503
- Marks DS, Colwell LJ, Sheridan R, Hopf TA, Pagnani A, Zecchina R, Sander C (2011) Protein 3D structure computed from evolutionary sequence variation. *PLoS ONE* 6(12):e28766. doi:10.1371/journal.pone.0028766
- Markwick PR, Cervantes CF, Abel BL, Komives EA, Blackledge M, McCammon JA (2010) Enhanced conformational space sampling improves the prediction of chemical shifts in proteins. *J Am Chem Soc* 132(4):1220–1221. doi:10.1021/ja909369z
- Meiler J (2003) PROSHIFT: protein chemical shift prediction using artificial neural networks. *J Biomol NMR* 26(1):25–37. doi:10.1007/s10858-002-2785-5

- Mittag T, Forman-Kay JD (2007) Atomic-level characterization of disordered protein ensembles. *Curr Opin Struct Biol* 17(1):3–14. doi:10.1016/j.sbi.2007.01.009
- Moseley HN, Sperl LJ, Rienstra CM (2010) Automated protein resonance assignments of magic angle spinning solid-state NMR spectra of beta1 immunoglobulin binding domain of protein G (GB1). *J Biomol NMR* 48(3):123–128. doi:10.1007/s10858-010-9448-2
- Mulder FA (2009) Leucine side-chain conformation and dynamics in proteins from ^{13}C NMR chemical shifts. *ChemBioChem* 10(9):1477–1479. doi:10.1002/cbic.200900086
- Neal S, Nip AM, Zhang H, Wishart DS (2003) Rapid and accurate calculation of protein ^1H , ^{13}C and ^{15}N chemical shifts. *J Biomol NMR* 26(3):215–240
- Nielsen JT, Eghbalnia HR, Nielsen NC (2012) Chemical shift prediction for protein structure calculation and quality assessment using an optimally parameterized force field. *Prog Nucl Magn Reson Spectrosc* 60:1–28. doi:10.1016/j.pnmrs.2011.05.002
- Osapay K, Case DA (1991) A new analysis of proton chemical shifts in proteins. *J Am Chem Soc* 113(25):9436–9444. doi:10.1021/ja00025a002
- Park SH, Das BB, Casagrande F, Tian Y, Nothnagel HJ, Chu M, Kiefer H, Maier K, De Angelis A, Marassi FM, Opella SJ (2012) Structure of the chemokine receptor CXCR1 in phospholipid bilayers. *Nature* (in press)
- Pastore A, Saudek V (1990) The relationship between chemical shift and secondary structure in proteins. *J Magn Reson* 90(1):165–176. doi:10.1016/0022-2364(90)90375-j
- Raman S, Lange OF, Rossi P, Tyka M, Wang X, Aramini J, Liu G, Ramelot TA, Eletsky A, Szyperski T, Kennedy MA, Prestegard J, Montelione GT, Baker D (2010) NMR structure determination for larger proteins using backbone-only data. *Science* 327(5968):1014–1018. doi:10.1126/science.1183649
- Saito H (1986) Conformation-dependent ^{13}C chemical shifts: a new means of conformational characterization as obtained by high-resolution solid-state ^{13}C NMR. *Magn Reson Chem* 24(10):835–852. doi:10.1002/mrc.1260241002
- Sharma M, Yi M, Dong H, Qin H, Peterson E, Busath DD, Zhou HX, Cross TA (2010) Insight into the mechanism of the influenza A proton channel from a structure in a lipid bilayer. *Science* 330(6003):509–512. doi:10.1126/science.1191750
- Shen Y, Bax A (2007) Protein backbone chemical shifts predicted from searching a database for torsion angle and sequence homology. *J Biomol NMR* 38(4):289–302. doi:10.1007/s10858-007-9166-6
- Shen Y, Bax A (2010) modest improvement in empirical NMR chemical shift prediction by means of an artificial neural network. *J Biomol NMR* 48(1):13–22. doi:10.1007/s10858-010-9433-9
- 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(12):4685–4690. doi:10.1073/pnas.0800256105
- Shen Y, Vernon R, Baker D, Bax A (2009) De novo protein structure generation from incomplete chemical shift assignments. *J Biomol NMR* 43(2):63–78. doi:10.1007/s10858-008-9288-5
- Spera S, Bax A (1991) Empirical correlation between protein backbone conformation and C. alpha. and C. beta. ^{13}C nuclear magnetic resonance chemical shifts. *J Am Chem Soc* 113(14):5490–5492. doi:10.1021/ja00014a071
- Steele RA, Opella SJ (1997) Structures of the reduced and mercury-bound forms of MerP, the periplasmic protein from the bacterial mercury detoxification system. *Biochemistry* 36(23):6885–6895. doi:10.1021/bi9631632
- Szilagy L, Jardetzky O (1989) Proton chemical shifts and secondary structure in proteins. *J Magn Reson* 83(3):441–449. doi:10.1016/0022-2364(89)90341-7
- Tian F, Valafar H, Prestegard JH (2001) A dipolar coupling based strategy for simultaneous resonance assignment and structure determination of protein backbones. *J Am Chem Soc* 123(47):11791–11796
- Tycko R, Hu KN (2010) A Monte Carlo/simulated annealing algorithm for sequential resonance assignment in solid state NMR of uniformly labeled proteins with magic-angle spinning. *J Magn Reson* 205(2):304–314. doi:10.1016/j.jmr.2010.05.013
- Vila JA, Villegas ME, Baldoni HA, Scheraga HA (2007) Predicting ^{13}C alpha chemical shifts for validation of protein structures. *J Biomol NMR* 38(3):221–235. doi:10.1007/s10858-007-9162-x
- Vila JA, Aramini JM, Rossi P, Kuzin A, Su M, Seetharaman J, Xiao R, Tong L, Montelione GT, Scheraga HA (2008) Quantum chemical ^{13}C (alpha) chemical shift calculations for protein NMR structure determination, refinement, and validation. *Proc Natl Acad Sci USA* 105(38):14389–14394. doi:10.1073/pnas.0807105105
- Vila JA, Arnautova YA, Martin OA, Scheraga HA (2009) Quantum-mechanics-derived ^{13}C alpha chemical shift server (CheShift) for protein structure validation. *Proc Natl Acad Sci USA* 106(40):16972–16977. doi:10.1073/pnas.0908833106
- Villegas ME, Vila JA, Scheraga HA (2007) Effects of side-chain orientation on the ^{13}C chemical shifts of antiparallel beta-sheet model peptides. *J Biomol NMR* 37(2):137–146. doi:10.1007/s10858-006-9118-6
- Wang Y, Jardetzky O (2004) Predicting ^{15}N chemical shifts in proteins using the preceding residue-specific individual shielding surfaces from phi, psi i-1, and chi 1 torsion angles. *J Biomol NMR* 28(4):327–340. doi:10.1023/B:JNMR.0000015397.82032.2a
- Wishart DS, Sykes BD, Richards FM (1991) Relationship between nuclear magnetic resonance chemical shift and protein secondary structure. *J Mol Biol* 222(2):311–333
- 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(Web Server issue):496–502. doi:10.1093/nar/gkn305
- Xu XP, Case DA (2001) Automated prediction of ^{15}N , ^{13}C alpha, ^{13}C beta and $^{13}\text{C}'$ chemical shifts in proteins using a density functional database. *J Biomol NMR* 21(4):321–333
- Yarov-Yarovoy V, Schonbrun J, Baker D (2006) Multipass membrane protein structure prediction using Rosetta. *Proteins* 62(4):1010–1025. doi:10.1002/prot.20817