Northeast Structural Genomics Consortium
Quantum-Mechanics-Derived $^{13}C^{\alpha}$ Chemical Shifts for Protein Structure Validation

A burgeoning repository of protein chemical shift data has become widely available to the scientific community in recent years (www.bmrb.wisc.edu). This database is quite valuable, since it has been long recognized that the observed chemical shift of a nucleus is highly sensitive to its chemical environment, and can be used directly in structure refinement protocols. In particular, the $^{13}C^{\alpha}$ chemical shift for a given amino acid is almost completely determined by its backbone and side chain torsional angles, and is not correlated with neighboring residues in the sequence. Scheraga and co-workers have demonstrated that for a given conformation, $^{13}C^{\alpha}$ chemical shifts for each amino acid in a protein can be computed at the density functional level of theory (DFT) by treating each as a terminally blocked tripeptide (Ac-GXG-NMe). Although computationally intensive, the approach offers a number of advantages including i) $^{13}C^{\alpha}$ chemical shifts can be calculated with high accuracy, ii) there is no need for *a priori* knowledge of the oligomerization state of the protein, and iii) no knowledge-based information is required. They developed a new, purely physics-based structure validation metric called the conformationally averaged root-mean-square-deviation, ca-rmsd (Vila *et al.*, 2007, J. Biomol. NMR, *38*, 221), which reflects the agreement between experimental and computed $^{13}C^{\alpha}$ chemical shifts for all residues in an ensemble of protein structures (Eq. 1),

$$ca-rmsd^{\alpha} = \left[(1/N) \sum_{\mu=1}^{N} (\Delta_{\mu}^{\alpha})^2 \right]^{1/2} \qquad (1)$$

where $1 \le \mu \le N$, with N being the number of observed $^{13}C^{\alpha}$ chemical shifts, and $\Delta_{\mu}^{\alpha} = (^{13}C_{obs,\mu} - \langle ^{13}C_{comp} \rangle_{\mu}$.

For a single conformer (i.e., an X-ray structure) ca-rmsd is equivalent to the $^{13}C^{\alpha}$ rmsd (Eq. 2).

$$ca-rmsd^{\alpha} = rmsd^{\alpha} = \left[(1/N) \sum_{\mu=1}^{N} (^{13}C^{\alpha}_{observed,\mu} - ^{13}C^{\alpha}_{computed,\mu})^2 \right]^{1/2} \qquad (2)$$

In the NESG project, this physics-based approach has been applied to validate the NMR and X-ray structures of the YnzC protein from *Bacillus subtilis*, where it was demonstrated that ensembles of structures derived from NMR data are a better representation of the $^{13}C^{\alpha}$ chemical shifts than the crystal structure (Vila *et al.*, 2008).

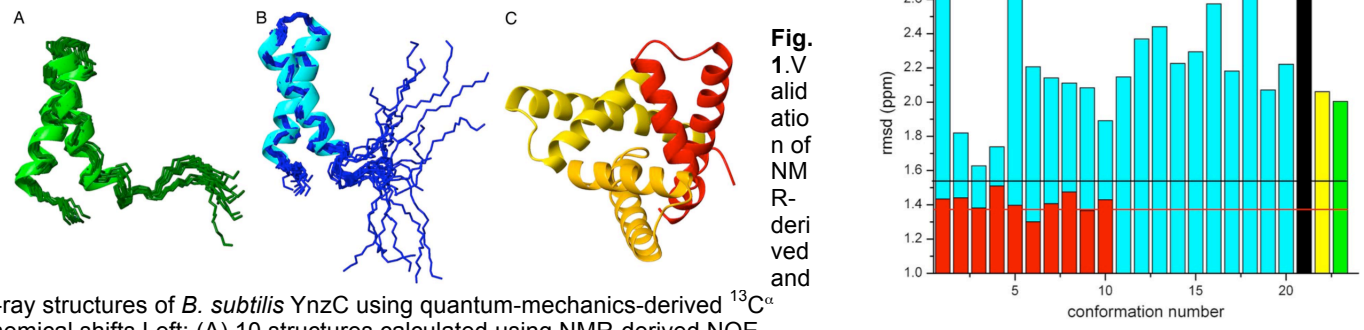

Fig. 1. Validation of NMR-derived and X-ray structures of *B. subtilis* YnzC using quantum-mechanics-derived $^{13}C^{\alpha}$ chemical shifts.Left: (A) 10 structures calculated using NMR-derived NOE constraints and torsional angles computed from $^{13}C^{\alpha}$ chemical shifts using DFT, (B) the final ensemble of 20 NMR structures (PDB entry, 2JVD), and (C) the 2.0 Å X-ray structure which features three monomers in the unit cell (PDB entry, 3BHP). Right: Plot of rmsd between computed and observed $^{13}C^{\alpha}$ chemical shifts for each of the conformers derived from A (red), B (cyan) and C (black, yellow, and green). Horizontal lines denote the ca-rmsds for the NMR-derived structures, A (red) and B (black).

Recently, Scheraga and co-workers introduced a new quantum-mechanics-derived CheShift server (http://cheshift.com) for computing $^{13}C^{\alpha}$ chemical shifts of proteins based on their structure (Vila *et al.*, 2009, Proc. Natl. Acad. Sci. USA, in press). Internally, the server consults a comprehensive library of $^{13}C^{\alpha}$ chemical shifts computed at the DFT level of theory using a smaller basis set (for increased speed) for all 20 naturally occurring amino acids from almost 700,000 conformations generated as a function of backbone (ϕ,ψ,ω) and side chain (χ_1,χ_2) torsional angles. The server is being used in NESG NMR projects to evaluate or rank protein models ("decoys") obtained in intermediate stages of automated protein NMR structure refinement and/or automated NOESY assignment process, and for structure validation.

Vila, J.; Aramini, J.M.; Rossi, P.; Kuzin, A.; Su, M.; Seetharaman, J.; Xiao, R.; Tong, L.; Montelione, G.T.; Scheraga, H. Proc. Natl. Acad. Sci. U.S.A. 2008, 105:14389 – 14394. Quantum chemical 13Calpha chemical shift calculations for Protein NMR structure determination, refinement and validation.