CheShift-2: graphic validation of protein structures

Osvaldo A. Martin1, Jorge A. Vila1,2,∗ and Harold A. Scheraga2,∗

1Universidad Nacional de San Luis, IMASL-CONICET, Ejército de Los Andes, 950-5700 San Luis, Argentina and 2Baker Laboratory of Chemistry and Chemical Biology, Cornell University, Ithaca, NY 14853-1301, USA

Associate Editor: Anna Tramontano

ABSTRACT
Summary: The differences between observed and predicted 13Cα chemical shifts can be used as a sensitive probe with which to detect possible local flaws in protein structures. For this reason, we previously introduced CheShift, a Web server for protein structure validation. Now, we present CheShift-2 in which a graphical user interface is implemented to render such local flaws easily visible. A series of applications to 15 ensembles of conformations illustrate the ability of CheShift-2 to locate the main structural flaws rapidly and accurately on a per-residue basis. Since accuracy plays a central role in CheShift predictions, the treatment of histidine (His) is investigated here by exploring which form of His should be used in CheShift-2.

Availability: CheShift-2 is free of charge for academic use and can be accessed from www.cheshift.com

Received on January 4, 2012; revised on March 27, 2012; accepted on March 31, 2012

1 INTRODUCTION

Chemical shifts provide important information about the conformations of proteins in solution (see, for example, Wishart, 2011, and references therein). For this reason, we developed CheShift-2, a Web server for protein structure validation based on a quantum mechanics database of 13Cα chemical shifts (Vila et al., 2009). CheShift was originally developed to return a list of predicted values of 13Cα chemical shifts. It was the user’s responsibility to compare the predicted with the observed 13Cα chemical shifts to assess the global quality of a protein. However, it is a highly desirable goal of any accurate validation method (Nabuurs et al., 2006; Vila and Scheraga, 2009) to identify the existence of local flaws, in addition to the global quality; see analysis of local versus global chemical shift validation of Dynein light chain 2A protein in Supplementary Material. In order to automate and facilitate the validation process on a per-residue basis, we added a GUI to CheShift. The GUI displays the differences between observed and predicted 13Cα chemical shifts by using a four-color code mapped onto a 3D protein model. A set of 15 proteins was used to test the ability of CheShift-2 to detect local flaws. This set was selected from the Protein Data Bank (PDB), Berman et al., 2000) and corresponds to observed and superseded NMR protein structures. Released PDB data (coordinates and experimental data) are rendered obsolete when the authors have collected new data or had re-refined the structure. The obsolete entry is usually replaced by a new (superseding) entry that receives a new PDB ID.

2 METHODS

For each amino acid μ, it is possible to define the difference between observed and predicted 13Cα chemical-shifts as:

where, 13Cα,observed,μ is the chemical shift of residue μ in conformation i out of Ω conformations. The average of the predicted chemical shifts over the Ω conformations is evaluated because proteins in solution exist as an ensemble of conformations.

The following procedure for mapping the Δμ values onto a 3D protein model was formulated. First, the Δμ value computed for each residue μ is smoothed by averaging it over the values of the two nearest-neighbor residues (see Supplementary Material for details). Second, the resulting averaged Δμ value is discretized according to the following rule:

The selection of the cut-off σ value of 1.7 ppm is explained in the Supplementary Material. Third, the Δμ values at on the 3D protein model and associated with a color; blue, white and red, respectively. Implicit in this color-code assignment is the assumption that average differences per residue between observed and predicted 13Cα chemical shifts that are within ∼1σ (blue) are considered small; within ∼2σ (white) they are considered medium, i.e. being both blue and white considered as acceptable differences; and beyond 2σ (red), they are considered large differences and, hence, special attention should be attached to those residues. In addition, the color yellow was adopted to indicate the absence of the observed or computed 13Cα chemical shift value.

3 RESULTS

We found evidence (see Supplementary Material) indicating that the protonated form of histidine (His), rather than the neutral ones, namely, the Nδ1-H or Nε2-H tautomers form, respectively, leads to a better representation of the observed 13Cα chemical shifts. This observation, together with the well-documented effect of proline on the computed chemical shift of the preceding residue (Vila et al., 2010), are now taken into account in CheShift-2 predictions.

Figure 1A shows the color distribution obtained for three ensembles of conformations for the bovine cytochrome B5 protein. The first ensemble of conformations was obtained by NMR
spectroscopy (PDB ID 1WDB); most of the flaws (red-colored residues) are located in the helices, which are very distorted. In the year 2003, 1WDB was superseded by 1HKO, also determined by NMR spectroscopy. According to CheShift-2, 1HKO is enriched in blue regions, indicating that it is indeed a better structural model than 1WDB. A third conformation (PDB ID 1CYO) is included for comparison with the previous two NMR-derived conformations. This third conformation was determined by X-ray diffraction at 1.5 Å resolution, is enriched in blue/white regions, confirming that the ICMI protein is indeed a very good structure [see Fig. 1B (III)].

4 CONCLUSIONS

CheShift-2 constitutes a fast and accurate validation tool with which to determine the existence of local flaws in protein models. Examples analyzed in the present study show that, if the NMR-determined ensemble had not been solved at a high-quality level, a comparison with the corresponding structure determined by X-ray crystallography reveals that the X-ray structure is almost flawless and, hence, indicates that the detected flaws in the NMR-determined ensemble are not a bias of the method but a warning that the NMR-derived structure may benefit from further structural refinement.

This new physics-based validation tool, CheShift-2, should be used as a complementary one to other existing knowledge-based methods, such as WHAT IF (Vriend, 1990) and PROCHECK (Laskowski et al., 1993), or combined knowledge-based and physics-based methods, such as the PSVS package (Huang et al., 2005; Bhattacharya et al., 2007).

ACKNOWLEDGEMENTS

We thank P. Serrano at the Scripps Research Institute for valuable discussions regarding protein structure validation methods. The research was conducted by using the resources of Pople, a facility of the NSF Terascale Computing System at the Pittsburgh Supercomputer Center.

Funding: NIH (GM-14312) and NSF (MCB10-19767), USA; CONICET and UNSL (P-328402), Argentina.

Conflict of Interest: none declared.

REFERENCES

Berman,H.M. et al. (2000) The Protein Data Bank. Nucleic Acids Res., 28, 335–342.

Bhattacharya,A. et al. (2007) Evaluating protein structures determined by structural genomics consortium. Proteins, 66, 778–795.

Huang,Y.L. et al. (2005) Protein NMR Recall, Precision and F-measure scores (RPF scores): structure quality assessment measures based on information retrieval statistics. J. Am. Chem. Soc., 127, 1665–1674.

Laskowski,R.A. et al. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. J. Appl. Cryst., 26, 283–291.

Nabutov,S.B. et al. (2006) Traditional biomolecular structure determination by NMR spectroscopy allows for major errors. PLoS Comp. Biol., 2, 71–79.

Vila,J.A. and Scheraga,H.A. (2005) Assessing the accuracy of protein structures by quantum mechanical computations of 13C chemical shifts. J. Mol. Biol., 348, 23–30.

Vila,J.A. et al. (2009) Quantum-mechanics-derived 13C chemical shift server (CheShift) for protein structure validation. Proc. Natl Acad. Sci., 106, 16972–16977.

Vila,J.A. et al. (2010) Sequential nearest-neighbor effects on computed 13C chemical shifts. J. Biomol. NMR, 48, 23–30.

Vriend,G. (1990) WHAT IF: a molecular modeling and drug design program. J. Mol. Graph., 8, 52–56.

Wishart,D.S. (2011) Interpreting protein chemical shift data. Prog. Nucl. Magn. Reson. Spectrosc., 58, 62–67.