Structural bioinformatics

NQ-Flipper: validation and correction of asparagine/glutamine amide rotamers in protein crystal structures

Christian X. Weichenberger and Manfred J. Sippl*
Center of Applied Molecular Engineering, University of Salzburg, Jakob Haringerstrasse 5, 5020 Salzburg, Austria

Received on January 12, 2006; revised on March 6, 2006; accepted on March 30, 2006
Advance Access publication April 4, 2006
Associate Editor: Anna Tramontano

ABSTRACT
Summary: The error rate of asparagine (Asn) and glutamine (Gln) amide rotamers in protein crystal structures is in the order of 20% and as a consequence the current Protein Database (PDB) contains approximately half a million incorrect Asn and Gln side-chain rotamers. Here we present NQ-Flipper, a web service based on knowledge-based potentials of mean force to automatically detect and correct erroneous rotamers. We achieve excellent agreement with expert curated data.

Availability: The program is accessible freely as a web service at http://flipper.services.came.sbg.ac.at
Contact: sippl@came.sbg.ac.at

1 INTRODUCTION

The side-chain amide groups of asparagine (Asn) and glutamine (Gln) act simultaneously as hydrogen bond donors and acceptors. The electron density near the nitrogen and oxygen atoms is frequently compatible with two rotamers which are related by a 2-fold symmetry axis. This hampers the correct interpretation of electron density maps resulting in the perpetual assignment of incorrect rotamers with an error rate of 20% (McDonald and Thornton, 1995; Word et al., 1999). Stated in this way the problem seems to be specific for X-ray analysis of protein crystals but we emphasize that a similar error rate of 23% is found in NMR structures.

Since Asn and Gln residues frequently participate in hydrogen bond networks and functional groups, incorrect rotamers may severely interfere with physico-chemical studies of protein structures and molecular modeling tasks. On the other hand, incorrect rotamers in general result in unfavorable interactions which should be clearly detectable by proper energy calculations.

Currently, two web-based services are available for the detection of incorrect Asn and Gln rotamers. Lovell et al. (2003) identify correct rotamers by minimizing steric clashes after adding hydrogen atoms to the protein structure. Hooft et al. (1996) optimize the hydrogen bonding network of proteins allowing Asn and Gln residues to flip during the optimization process. Both methods use artificially created hydrogen atom positions for rotamer characterization. The service provided by Lovell et al. (2003) offers visualization and download of corrected PDB (Berman et al., 2000) entries.

2 IMPLEMENTATION AND USAGE

To address the rotamer problem we use potentials of mean force for pairwise interactions among all heavy atoms of standard amino acids (Sippl, 1990). For each Asn and Gln residue we compute the energy $\epsilon(R_1)$ of the original conformation $R_1$ as found in the PDB structure and the energy $\epsilon(R_2)$ for the alternative rotamer $R_2$. The energy difference $\Delta \epsilon := \epsilon(R_1) - \epsilon(R_2)$ serves as a score and from Boltzmann’s distribution we obtain the probabilities $p(R_1) = 1/(1 + \exp(\Delta \epsilon))^{-1}$ and $p(R_2) = 1 - p(R_1)$. Given a threshold value $\nu$ (see below), we distinguish three cases. If $\Delta \epsilon < -\nu$, the probability $p(R_1)$ is close to one and $R_1$ is considered to be the correct rotamer. If $\Delta \epsilon > \nu$ then the probability $p(R_2)$ is close to zero and $R_1$ is the incorrect rotamer. For $|\Delta \epsilon| < \nu$, both rotamers have significant probabilities so that they are likely to coexist in the crystal structure and the assignment is ambiguous. In particular, for $\Delta \epsilon = 0$ both rotamers have equal probabilities $p(R_1) = p(R_2) = 1/2$.

A suitable value of $\nu$ is obtained by comparison with the reference set reported by Word et al. (1999). The reference set consists of 100 protein chains containing 1006 (75.9%) Asn and Gln residues classified as correct and 320 (24.1%) classified as incorrect. Using established terminology (Baldi et al., 2000) we find for incorrect rotamers defined as $\Delta \epsilon > \nu = 6$ a sensitivity of 92.7%, a specificity of 96.7% and an overall accuracy of 95.8%. The fraction of ambiguous rotamers, defined by $-6 \leq \Delta \epsilon \leq 6$ is 5.8%.

The mean force potentials are first compiled from a database containing the original $R_1$ rotamers. The potentials are then refined by several cycles of rotamer correction and recompilation of potentials. The potentials converge quickly to a stable self-consistent solution. A subsequent comparison shows excellent agreement with rotamer flips independently suggested by expert analysis [e.g. Word et al. (1999)]. In principle the approach presented here can be applied to the related problem of ambiguities in His conformers. However, the respective analysis requires a careful consideration of various protonation states which is beyond the scope of the present analysis.

As an example we provide an analysis of oxidoreductase 1ra9 (resolution 1.55 Å). The structure contains four residues with significant $\Delta \epsilon$-scores (Asn-18, Asn-23, Gln-65, Gln-108; $\Delta \epsilon$-scores 20.5, 16.7, 12.8 and 26.7, respectively). These residues should be flipped to the alternative rotamer $R_2$ which is corroborated by a detailed analysis of the interactions among the affected atoms (Fig. 1).
The method presented here is implemented as a web service called NQ-Flipper (http://flipper.services.came.sbg.ac.at). The service provides validation and correction of Asn and Gln residues in protein structures that can be specified either as a valid PDB code or uploaded as a PDB formatted file. For each Asn and Gln amino acid in the structure the server computes ΔΦ scores by taking into account all chains and the full crystal symmetry.

The results are presented in the form of a table of ΔΦ values and a graphical view of the structure based on Jmol. The table signifies incorrect residues in red and residues which are within a radius of 8 Å of non-standard groups in blue. The residues in blue have to be treated with caution since they may interact with atoms whose potentials of mean force are currently not available. All assignments can be edited by the user and the corrected coordinate files can be downloaded in various compression formats.

All transactions are encrypted by the https protocol and the data are stored in session dependent directories that are only accessible to the user who has control of the session. A detailed description of all parameters is available in the help section of the website.

ACKNOWLEDGEMENTS

The authors thank Ralf Grosse-Kunstleve for kind permission to use his sglite crystallographic symmetry library.

Conflict of Interest. none declared.

REFERENCES