NQ-Flipper: recognition and correction of erroneous asparagine and glutamine side-chain rotamers in protein structures

Christian X. Weichenberger and Manfred J. Sippl*

Center of Applied Molecular Engineering, University of Salzburg, Hellbrunnerstrasse 34, 5020 Salzburg, Austria

Received January 31, 2007; Revised March 30, 2007; Accepted April 8, 2007

ABSTRACT

The current Protein Data Bank (PDB) contains about 40 000 protein structures with approximately half a million incorrect atom positions resulting from erroneously assigned asparagine (Asn) and glutamine (Gln) rotamers. These errors affect applications in protein structure analysis, modeling and docking and therefore the detection, correction and prevention of such errors is highly desirable. We present NQ-Flipper, a web service based on mean force potentials to automatically detect and correct erroneous Asn and Gln rotamers. The service accepts protein structure files formatted in PDB style or PDB codes. For an Asn/Gln side-chain amide NQ-Flipper computes the total interaction energy with the surrounding atoms as the sum of pairwise atom–atom interaction energies. The energy difference between the original and the alternative rotamers identifies the correct configuration of the amide group. The web service lists the interaction energies of all Asn/Gln residues found in a PDB file and shows the structure and offending residues in an interactive 3D viewer. The corrected protein structure is available for download in various compression formats. The web service is accessible at http://flipper.services.came.sbg.ac.at

INTRODUCTION

The side-chains of the amino acids asparagine (Asn) and glutamine (Gln) terminate by an amide group. The amide group may form several hydrogen bonds with the surrounding atoms and thus frequently plays an important role in protein structure stability and in intermolecular interactions (1–8). A single amide group may form up to four hydrogen bonds, two donated by the nitrogen and two accepted by the oxygen. Hence, an incorrect configuration generally results in highly unfavorable interaction energies and in view of the refinement protocols used in X-ray analysis, which frequently employ energy calculations, the occurrence of errors may seem unlikely. Nevertheless, the average error rate found in the current Protein Data Bank (PDB) (9) is of the order of 20% (10–14). The high error rate arises from limitations encountered in X-ray analysis. The amide nitrogen and oxygen atoms have similar electron densities and are indistinguishable in electron density maps at moderate and low resolutions and it is generally thought that structures of higher resolution (1.5 Å or better) are largely error free. From this point of view the Asn/Gln rotamer problem seems to be a peculiarity of X-ray analysis. However, a similar error rate is found for solution structures of proteins determined by NMR (10).

Presently two web services provide information on Asn/Gln rotamers. The PDBREPORT database/WHATIF server (15) presents a static view on a plethora of automatically generated quality scores. Among these is a listing of incorrect Asn/Gln rotamers. The WHATIF server displays information but does not provide a mechanism to correct erroneous Asn/Gln rotamers. The MolProbity (16) server is dedicated to protein structure analysis and offers a variety of interactive tools to validate either an uploaded coordinate file or a structure from PDB. One special subtask is the validation and correction of Asn/Gln rotamers by adding hydrogens and picking the rotamer with lesser steric clashes.

Here we present NQ-Flipper, a web service for interactive validation and correction of Asn/Gln amide rotamers. The server operates in three steps. A protein coordinate file in PDB format is uploaded and scores based on potentials of mean force (10, 17) are computed. In the second step, the results are displayed in a table where incorrect Asn/Gln rotamers are marked and the respective structure is shown in an interactive 3D viewer. The corrected protein structure is available for download in various compression formats. The web service is accessible at http://flipper.services.came.sbg.ac.at

*To whom correspondence should be addressed. Tel: +43 662 8044 5797; Fax: +43 662 8044 176; Email: sippl@came.sbg.ac.at

© 2007 The Author(s)
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/2.0/uk/) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.
structures. Rotamers are reported with their associated Δɛ-score and erroneous rotamers are replaced by their corrected conformations resulting in a separate coordinate file available for download. NQ-Flipper has a small set of parameters that may be controlled and optionally changed by the user. Protein structures are entered at the main page along with a small number of parameters as described in the following paragraphs.

PDB code
A valid PDB four letter code specifies the protein structure to be processed. The repository of coordinate files maintained by NQ-Flipper is concurrent with PDB (9).

File name
Any coordinate file compliant with the PDB file format can be uploaded. When determining a new structure this allows rotamer validation at any stage of the protein crystal structure refinement process. Crystallographic symmetry is used to generate the complete crystal structures so that complete structures are used for rotamer assignment. Valid input file compression formats include gzip and unix compress.

Model number
Computations for coordinate files containing more than one model are restricted to a single model identified by this number.

Altloc indicator
Treatment of alternate location (altloc) indicators is similar to model numbers in that coordinates are retrieved for a certain altloc character only. The usage of altloc indicators is not clearly defined by PDB. In extreme cases all possible combinations of atoms with alternate locations may have to be considered. Depending on the particular PDB file this may result in an enormous number of possible combinations. A consistent treatment of alternate locations requires the submission of a complete model for each variant.

Threshold
A threshold ɛ applied to Δɛ-scores is used to distinguish rotamers with single and multiple occupations. The larger the value of ɛ the larger is the number of Asn/Gln amides considered to occupy both rotamers. The NQ-Flipper web page provides a statistical analysis of the agreement of NQ-Flipper results with a refined reference data set (12) as a function of the threshold ɛ.

Rotate/Flip
Coordinates for the alternative rotamer atoms are derived by a 180° rotation about the Cα – Cγ (Asn) and Cγ – Cδ (Gln) axis. This preserves bond angles and distances which is not guaranteed for the ‘Flip’ option where only atom identities are swapped. The latter option has been applied to correct Asn/Gln rotamers in the reference set (12). The option is offered here for comparison and the computation of the statistical analysis, but its use correcting Asn/Gln...
rotamers is strongly discouraged, since it results in improper covalent geometry of the amide atoms.

For a moderately sized protein the Asn/Gln rotamers are validated within seconds. The results are presented in tabular form listing all Asn/Gln residues sorted and color-coded by their associated Δε-score. Entries highlighted in red refer to rotamers with unfavorable interaction energies. An interactive 3D-viewer based on Jmol (http://www.jmol.org) displays the Cα backbone trace of the protein structure with side-chain atoms

![Figure 1. NQ-Flipper analysis of Asn/Gln rotamers.](image)

Each residue is listed with its Δε-score and associated Δε-value as displayed on the NQ-Flipper web page. Coordinates were derived for alternate location indicator A. The software detects three Asn residues with unfavorable Δε-scores, i.e., rotamers with Δε > v where the threshold value is set to v = 6 (the default). These incorrect rotamers are highlighted in red and include Asn-1114 (∆ε = +43.3), Asn-1088 (∆ε = +39.9) and Asn-1020 (∆ε = +12.3). One residue, Gln-1133 (∆ε = 5.1), is considered ‘ambiguous’ since its Δε-score is in the range where both rotamers are occupied with non-negligible probability (|Δε| < 6). Such residues of multiple occupancy are colored orange. All remaining Asn/Gln residues have Δε-scores below –6 indicating correct rotamers. Residue Asn-1016 (∆ε = –24.0) shown in blue is in the vicinity of an isopropanol buffer molecule, IPA. The minimum distance between atoms from the IPA molecule and the respective amide group in Asn-1016 is 5.7 Å. The residue is colored blue to indicate that it is close to a heterogeneous group so that some important interactions may not be included in the score—although even in such cases a score greater than v generally indicates an incorrect rotamer. Figure 1c and d provide a detailed view of two of the incorrect rotamers in the conformation $R_1$, as found in the PDB file, interacting with their respective chemical environment. The atoms are colored by atom type: carbon, gray; nitrogen, blue; oxygen, red. The dashed lines indicate distances in Å between atoms. (c) In the original conformation residue Asn-1020 (∆ε = +12.3) is unable to participate in a hydrogen bond network. In the alternative conformation $R_2$ the amide nitrogen and oxygen atoms swap positions enabling hydrogen bond formation between the amide nitrogen and the carbonyl oxygen of Asn-1016, the amide oxygen and the side-chain nitrogen of Lys-1015 and a nitrogen of the guanidinium group of Arg-1118. Note that by flipping the amide group several ‘anti-hydrogen bonds’ of very high energy are replaced by genuine hydrogen bonds of very low energy. (d) Residue Asn-1088 (∆ε = +39.9), where the amide nitrogen and oxygen atoms have unfavorable interactions with the respective main chain nitrogen and oxygen atoms of Ser-1074. The amide oxygen of this residue additionally interacts unfavorably with Gln-1093, a correct rotamer since it forms a hydrogen bond to the carbonyl group of Gly-1086. Flipping Asn-1088 to rotamer $R_2$ results in the formation of three hydrogen bonds. Figure 1c and d were generated using PyMOL (http://pymol.sourceforge.net).
of incorrect rotamers rendered as spheres. Multiple occupancies (|Δε| < v) are indicated in orange and are left in their original conformation. Residues with side-chain amides closer than 8 Å to atom centers of non-standard groups are colored in blue. These are frequently residues participating in functional sites and therefore, they may be particularly important. The assignment produced by NQ-Flipper may be edited by the user. Corrected coordinate files can be downloaded in various compression formats. Figure 1 provides an example of autohydrolysin, PDB code 1mc2 (alternate location indicator A) (21), a phospholipase A2 at a resolution of 0.85 Å directly solved by dual-space Shake-and-Bake refinement. The structure contains eleven Asn/Gln residues. Out of these NQ-Flipper flags three rotamers as erroneous based on disagreement with empirical statistics which is also in line with basic physico-chemical principles.

CONCLUSION

The NQ-Flipper web service provides an interactive tool for the detection and correction of unfavorable Asn/Gln rotamers utilizing knowledge-based potentials of mean force derived from high resolution protein structures. Except for very large crystal structures the response time of the server is immediate (i.e. within seconds). The force derived from high resolution protein structures.

REFERENCES

1. Yoder, M.D., Keen, N.T. and Jurnak, F. (1993) New domain motif: the structure of perate lyase C, a secreted plant virulence factor. Science, 260, 1503–1507.
2. Battiste, J.L., Mao, H.Y., Rao, N.S., Tan, R.Y., Muhandiram, D.R., Kay, L.E., Frankel, A.D. and Williamson, J.R. (1996) Alpha helix-RNA major groove recognition in an HIV-1 Rev peptide RRE RNA complex. Science, 273, 1547–1551.
3. Coleman, D.E., Berghuis, A.M., Lee, E., Linder, M.E., Gilman, A.G. and Sprang, S.R. (1994) Structures of active conformations of Gi alpha 1 and the mechanism of GTP hydrolysis. Science, 265, 1405–1412.
4. Vernet, T., Tessier, D.C., Chatellier, J., Plouffe, C., Lee, T.S., Thomas, D.Y., Storer, A.C. and Menard, R. (1995) Structural and functional roles of asparagine 175 in the cysteine protease papain. J. Biol. Chem., 270, 16645–16652.
5. Raymond, A.C., Rideout, M.C., Staker, B., Hjerrild, K. and Burgin, A.B. (2004) Analysis of human tyrosyl-DNA phosphodiesterase I catalytic residues. J. Mol. Biol., 338, 895–906.
6. Fahn, S., Hileman, R.E., Fromm, J.R., Linhardt, R.J. and Rees, D.C. (1996) Heparin structure and interactions with basic fibroblast growth factor. Science, 271, 1116–1120.
7. Evdokimov, A.G., Anderson, D.E., Routzahn, K.M. and Waugh, D.S. (2001) Structural basis for oligosaccharide recognition by Pyrococcus furiosus maltodextrin-binding protein. J. Mol. Biol., 305, 891–904.
8. Schafer, K., Magnusson, U., Scheffel, F., Schiefer, A., Sandgren, M.O., Diederichs, K., Welte, W., Hulsmann, A., Schneider, E. et al. (2004) X-ray structures of the maltose-maltodextrin-binding protein of the thermoacidophilic bacterium alkaloclobacillus acidocaldarius provide insight into acid stability of proteins. J. Mol. Biol., 335, 261–274.
9. Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N. and Bourne, P.E. (2000) The Protein Data Bank. Nucleic Acids Res., 28, 235–242.
10. Weichenberger, C.X. and Sippl, M.J. (2000) Self-consistent assignment of asparagine and glutamine amide rotamers in protein crystal structures. Structure, 14, 967–972.
11. Weichenberger, C.X. and Sippl, M.J. (2006) NQ-Flipper: validation and correction of asparagine/glutamine amide rotamers in protein crystal structures. Bioinformatics, 22, 1397–1398.
12. Word, J.M., Lovell, S.C., Richardson, D.S. and Richardson, D.C. (1999) Asparagine and glutamine: using hydrogen atom contacts in the choice of side-chain amide orientation. J. Mol. Biol., 285, 1735–1747.
13. McDonald, J.K. and Thornton, J.M. (1995) The application of hydrogen bonding analysis in X-ray crystallography to help orientate asparagine, glutamine and histidine side chains. Protein Eng., 8, 217–224.
14. Hooff, R.W., Sander, C. and Vriend, G. (1996) Positioning hydrogen atoms by optimizing hydrogen-bond networks in protein structures. Proteins, 26, 363–376.
15. Hooff, R.W., Vriend, G., Sander, C. and Abola, E.E. (1996) Errors in protein structures. Nature, 381, 272.
16. Davis, I.W., Murray, L.W., Richardson, J.S. and Richardson, D.C. (2004) MOLPROBITY: structure validation and all-atom contact analysis for nucleic acids and their complexes. Nucleic Acids Res., 32, W615–W619.
17. Sippl, M.J. (1990) Calculation of conformational ensembles from potentials of mean force. An approach to the knowledge-based approach. J. Mol. Biol., 213, 859–883.
18. Chandler, D. (1987) Introduction to Modern Statistical Mechanics. Oxford University Press, New York.
19. Sippl, M.J. (1996) Helmholtz free energy of peptide hydrogen bonds in proteins. J. Mol. Biol., 260, 644–648.
20. Sippl, M.J., Ortner, M., Jaritz, M., Lackner, P. and Flockner, H. (1996) Helmholtz free energies of atom pair interactions in proteins.Fold. Des., 1, 289–298.
21. Liu, Q., Huang, Q., Teng, M., Weeks, C.M., Jelsch, C., Zhang, R. and Niu, L. (2003) The crystal structure of a novel, inactive, lysine 49 PL2 from Agkistrodon acutus venom: an ultra high resolution, ab initio structure determination. J. Biol. Chem., 278, 41400–41408.