Looking into the relationships between the quality of homology models and the accuracy of ligand-protein docking results

Bordogna A[1], Pandini A[2], Bonati L[1]

[1] DISAT, Università degli Studi di Milano-Bicocca, Milano, Italy
[2] Division of Mathematical Biology, National Institute for Medical Research, London, UK

Motivation

For an increasing number of cases, molecular docking calculations require the employment of three-dimensional models derived from homology modelling techniques. Therefore predicting the accuracy of the docking results on the basis of an a priori evaluation of the model quality is of great interest for a number of applications. Aim of this work is to identify quantitative relationships between indexes of model quality and the accuracy of binding geometries obtained by molecular docking calculations. The analysis is performed on a large test set assembled from a reference group of X-ray structures and a sample of corresponding homology models.

Methods

A representative subset of the CCDC/Astex Test Set [1] was chosen as the reference set of X-ray structures. This is a large and diverse set of known protein-ligand complexes extensively used to validate docking methods. In order to generate a corresponding dataset of models (266 entries), identification of candidate templates was performed by sequence similarity search using PSI-BLAST with default parameters till the convergence was reached [2]. Templates covering a broad range of sequence identity were selected to provide a reliable sampling of different evolutionary distances. To provide also a wide spectrum of quality, modelling was performed by employment of two methods: an automatic server, I-TASSER [3], and Modeller 9v1 [4]. Moreover, the reference alignments for Modeller were generated by using different approaches: T-Coffee [5] for mono- and multi-template sequence alignments; Praline [6] for profile-profile alignments; TMalign [7] for structural alignments. In a first step, the quality of the models was assessed by means of indexes derived from direct comparison to the known native structures: three structural alignment approaches implementing both global- and local-fit algorithms were employed (Dali, LGA, ProFit) [8,9,10]. These “calculated” indexes provided a direct measure of conformity to the target. Furthermore, a group of predictive indexes of model quality derived without any reference to the known native structures were employed (“predicted” indexes). A set of these was derived from a geometrical analysis with routine tools for structure validation [11,12,13] while another set is related to the sequence and structure similarity with the templates. AutoDock [14] was used to set up and perform docking calculations. The ligands were treated as flexible, whereas the proteins were treated as rigid bodies during docking simulations. In order to evaluate docking results, the dRMSD (distance Root Mean Square Deviation) was used. This is the root mean square deviation between the model ligand-site distances and the X-ray ligand-site corresponding distances.

Results

A first aim of this work was to investigate the correlations between model quality indexes calculated by direct comparison with the native structure and the docking quality index. For the broad and diverse test set here analysed, all these indexes showed a fairly good and statistically significant correlation with the dRMSD values of the best docking results. In particular, our analysis demonstrated that the quality of binding geometries obtained by molecular docking are mostly dependent on the ability of the model to correctly reproduce the active site geometry as well as on the relative structural distance of the model from the native structure compared to the template structure. The most ambitious aim was to predict also the quality of docking results from those indexes of model quality commonly used to validate both protein structures and theoretical models by comparison with background distributions or reference structures. Indeed, we showed that some of the “predicted” indexes selected for this analysis do exhibit a clear relationship with docking results' accuracy. In conclusion, for the first time, we provided a quantitative relationships between “calculated” indexes and docking results and our test demonstrated that further analyses on “predicted” indexes could lead to the identifica-

tion of general criteria for an a priori prediction of the accuracy in ligand-protein experiments.

References

[1)] Nissink, J. W. et al. (2002) Proteins, 49, 457-471.
[2)] Altschul, S.F. et al. (1997) Nucleic Acids Res., 25, 3389-3402.
[3)] Zhang, Y. (2008) BMC Bioinformatics, 9, 40.
[4)] Šali, A. et al. (1993) J. Mol. Biol., 234, 779-815.
[5)] Notredame, C. et al. (2000) J. Mol. Biol., 302, 205-217.
[6)] Heringa, J. (1999) Comput. Chem., 23, 341-364.
[7)] Zhang, Y. and Skolnick, J. (2005) Nucleic Acids Res., 33, 2302-2309.
[8)] Holm, L. and Sander, C. (1993) J. Mol. Biol., 233, 123-138.
[9)] Zemla, A. (2003) Nucleic Acids Res., 31, 3370-3374.
[10)] McLachlan, A. D. (1982) , Acta Cryst., A38, 871-873.
[11)] Bhattacharya, A. et al. (2007) Proteins, 66, 778-795.
[12)] Wallner, B. and Elofsson, A. (2003) Protein Sci., 12, 1073-1086.
[13)] McGuffin, L. J. (2008) Bioinformatics, 24, 586-587.
[14)] Huey, R. et al. (2006) J. Comput. Chem., 28, 1145-1152.

Contact : a.bordogna3@campus.unimib.it