BIOINFORMATICS APPLICATIONS NOTE Vol. 26 no. 13 2010, pages 1673–1674
doi:10.1093/bioinformatics/btq237

Structural bioinformatics Advance Access publication May 5, 2010

CHOYCE: a web server for constrained homology modelling with cryoEM maps

Reda Rawi[1,2,*], Lee Whitmore[1] and Maya Topf[1,*]
[1]Institute of Structural and Molecular Biology, Crystallography, Department of Biological Sciences, Birkbeck College, University of London, London, UK and [2]Department of Bioinformatics, Center for Medical Biotechnology, University of Duisburg-Essen, Universitaetsstr. 1-5, 45141, Essen, Germany
Associate Editor: Anna Tramontano

ABSTRACT

Summary: CHOYCE is a web server for homology modelling of protein components and the fitting of those components into cryo electron microscopy (cryoEM) maps of their assemblies. It provides an interactive approach to improving the selection of models based on the quality of their fit into the EM map.

Availability: http://choyce.ismb.lon.ac.uk/

Contact: m.topf@cryst.bbk.ac.uk; reda.rawi@uni-due.de

Supplementary information: Supplementary data are available at *Bioinformatics* online.

Received on December 4, 2009; revised on April 1, 2010; accepted on April 23. 2010

1 INTRODUCTION

In recent years, single-particle cryo electron microscopy (cryoEM) has become a major technique for structure determination of macromolecular assemblies (Frank, 2006). However, the majority of cryoEM density maps are still in the range of 5–15 Å, far from atomic resolution. Superior interpretation of these maps is now routinely achieved by fitting into them atomic structures of individual components, such as proteins, domains and even secondary structure elements (Chiu *et al*., 2005; Fabiola and Chapman, 2005). Many fitting methods have been developed, most of which rely on the cross-correlation function (CCF) between the component structure and the density map as a measure of goodness-of-fit (Fabiola and Chapman, 2005).

As experimental atomic structures of the assembly components are often not available, it is becoming increasingly common to fit comparative models (or homology models) instead (Topf and Sali, 2005). We recently showed that a correlation exists between the quality of a model [e.g. in terms of the root mean square deviation (RMSD) from the native structure] and its CCF with the density map (Topf *et al*., 2005). Based on that, a simple way to improve fitting accuracy is to generate a set of models using alternative templates and alignments. One of the most accurate models can be picked from this set based on its fit into the density and possibly in combination with other model assessment scores, such as statistical potentials of mean force (Shen and Sali, 2006; Topf *et al*., 2005, 2006). Unfortunately, researchers frequently do not use this option, despite the fact that it may lead not only to improved models, but also to a better analysis of the EM map.

Here we present CHOYCE, a web-based tool designed to perform homology modelling constrained by cryoEM density fitting, which enables the selection of improved models of assembly components. The CHOYCE server is freely available via the URL: http://choyce.ismb.lon.ac.uk/ (a valid licence key is required for the use of the MODELLER software).

2 METHODS

2.1 CHOYCE input

The server requires a density map and its parameters including the origin (in respect to the Cartesian coordinates) and the resolution (Å). The density map format can be either CCP4/MRC or XPLOR and its size must be smaller than 2 GB. It provides two input options: (i) an aligned pair of a target sequence and a template structure and (ii) an unaligned pair. In both cases, the template has to be pre-fitted in the density map, or at least placed in an approximated position.

2.2 Homology modelling

In the first step, CHOYCE generates homology models for the target sequence using the *automodel* class of the MODELLER-9v7 python interface (Eswar *et al*., 2008; Sali and Blundell, 1993). If the first input option is selected, CHOYCE uses the given target–template alignment to calculate n models, as requested by the user. If the second input option is selected, it generates l different alignments (as requested by the user) using the *salign* method (*align* or *align2d*) in MODELLER-9v7 (Madhusudhan *et al*., 2009), including one optimal alignment, and $l-1$ suboptimal alignments (Saqi and Sternberg, 1991). Ten models are then built per alignment ($n = 10 \times l$) and each of them is assessed by the statistical potential score DOPE (*normalized_dope* method in MODELLER-9v7) (Shen and Sali, 2006).

2.3 Density fitting and model assessment

In the next step, CHOYCE selects the model with the lowest DOPE score, and starting from the initial position for the template fitted into the input density map, it fits the model *locally* within the map. The local fitting is performed using the Scanning Monte Carlo (SMC) optimization option of Mod-EM, (*density.grid_search* in MODELLER-9v7, Topf *et al*., 2005), with the 'SPECIFIC' and 'GAUSS' options for the starting position and the atomic density function, respectively. Next, all the remaining models are superposed onto the fitted model using rigid-body least-squares minimization, as implemented in the *model.superpose* command of MODELLER-9v7. The user can also choose not to perform SMC, in which case all the models will be superposed onto the fitted template structure. Each of the models is then scored in the map based on their CCF with the map (Topf *et al*., 2005).

A typical CHOYCE job of modelling and fitting 50 models of 200 amino acids is completed in <10 min (independent of the size of the map). The computational cost scales linearly with the size of the protein and the number of models generated.

*To whom correspondence should be addressed.

© The Author(s) 2010. Published by Oxford University Press.
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/2.5), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

CHOYCE results appear in a web-based table, which may be accessed via a hyperlink automatically e-mailed to the user when the job is completed. The table consists of the names of *n* homology models built for the input target sequence and fitted in the EM map, sorted based on their CCF score (the highest score at the top). For each model both the CCF score and the normalized-DOPE score are reported. The coordinates of each model can be downloaded in a PDB format.

3 CASE STUDIES

As test cases, we used CHOYCE to generate two comparative models, one constrained by a simulated density map and the other by an experimental cryoEM map. In both cases, to measure the accuracy of the models, we calculated Cα RMSDs from the known native structures using least-square minimizations (*model.superpose* in MODELLER-9v7).

The first model is of subunit III of bovine heart cytochrome c oxidase, based on the structure of *E.coli* ubiquinol oxidase (PDB id: 1FFT, chain C, 22% sequence identity). The density map was generated from the native structure of cytochrome c oxidase (PDB id: 1OCC, chain C, residues 74–261) using the *molmap* command in Chimera (Pettersen *et al.*, 2004). We then used CHOYCE option 2/*align2d* to generate 50 models based on five alignments to 1FFT, and scored them applying the local fit first.

The second model is of the *S.cerevisiae* receptor of activated protein kinase C 1 (RACK1) (PDB id: 3FRX, chain A, residues 5-317), based on the structure of its homolog from *Arabidopsis thaliana* (PDB id: 3DM0, chain A, 49% sequence identity). For fitting, we used the 8.9 Åresolution cryoEM map of the 80S ribosome from the thermophilic fungus *T.lanuginosus* (EMDB ID: 1344, >85% sequence identity to *S.cerevisiae*). Because the density associated with RACK1 was identified on the small subunit (40S) where it is bound near the mRNA exit, we used only the density of the small subunit, which was previously segmented from the entire 80S density (Taylor *et al.*, 2009). We then fitted 3DM0 into the associated density and used CHOYCE option 2/*align* to generate and score 100 models based on 10 alignments to 3DM0, and scored them in the position of the fitted template.

In both test cases, the best-fitted model is the most accurate model (ranked first among 50 and 100 models, respectively). For 1OCC, the best model has Cα RMSD of 3.00 Å from the native structure. It is 18.5% more accurate than the best model based on the normalized-DOPE score (Cα RMSD = 3.68 Å, ranked last, Fig. 1). For RACK1, since the range of accuracy of the models was quite small (Cα RMSD between 2.84 and 3.07 Å), the best-fitted model is only slightly more accurate than the best normalized-DOPE model (2.84 Å versus 2.94 Å), although ranked much higher (1 versus 40) (Supplementary Fig. 1).

4 CONCLUSION

The CHOYCE web server provides a platform for homology modelling and fitting into cryoEM maps. The server aids in selecting one of the most accurate models from an ensemble of models based on the goodness-of-fit to the EM map. Future improvements will include modelling based on multiple template structures, as well as iterative alignment, fitting and refinement using a combination of CCF and statistical potential scores (Topf *et al.*, 2005, 2006).

ACKNOWLEDGEMENTS

We thank Andrej Sali, Ben Webb, Mallur Madhusudhan, David Houldershaw and Derek Taylor.

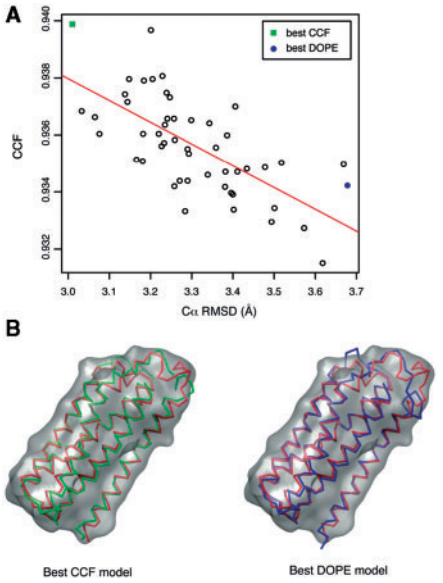

Fig. 1. Fitting of models into the simulated density map of 1OCC (10 Å resolution) using CHOYCE. (**A**) The CCF of the best fit between a model and the map is plotted against the Cα RMSD ($R^2 < 0.5$). The green solid square corresponds to the model with the best CCF and the blue solid dot corresponds to the model with the best normalized-DOPE score. (**B**) The native structure (red), the model based on the highest CCF (green, left), and the model based on the best DOPE score (blue, right) fitted in the map. Chimera (Pettersen *et al.*, 2004) was used for visualization.

Funding: Medical Research Council (G0600084) and Human Frontier Science Programme (RGY0079/2009-C) grants (to Topf lab).

Conflict of Interest: none declared.

REFERENCES

Chiu,W. *et al.* (2005) Electron cryomicroscopy of biological machines at subnanometer resolution. *Structure*, **13**, 363–372.

Eswar,N. *et al.* (2008) Protein structure modeling with MODELLER. *Methods Mol. Biol.*, **426**, 145–159.

Fabiola,F. and Chapman,M.S. (2005) Fitting of high-resolution structures into electron microscopy reconstruction images. *Structure*, **13**, 389–400.

Frank,J. (2006) *Three-Dimensional Electron Microscopy of Macromolecular Assemblies: Visualization of Biological Molecules in Their Native State*. Oxford University Press, New York, USA.

Madhusudhan,M.S. *et al.* (2009) Alignment of multiple protein structures based on sequence and structure features. *Protein Eng. Des. Sel.*, **22**, 569–574.

Pettersen,E.F. *et al.* (2004) UCSF Chimera–a visualization system for exploratory research and analysis. *J. Comput. Chem.*, **13**, 1605–1612.

Sali,A. and Blundell,T.L. (1993) Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.*, **234**, 779–815.

Saqi,M.A. and Sternberg,M.J. (1991) A simple method to generate non-trivial alternate alignments of protein sequences. *J. Mol. Biol.*, **219**, 727–732.

Shen,M.Y. and Sali,A. (2006) Statistical potential for assessment and prediction of protein structures. *Protein Sci.*, **15**, 2507–2524.

Taylor,D.J. *et al.* (2009) Comprehensive molecular structure of the eukaryotic ribosome. *Structure*, **17**, 1591–1604.

Topf,M. and Sali,A. (2005) Combining electron microscopy and comparative protein structure modeling. *Curr. Opin. Struct. Biol.*, **15**, 578–585.

Topf,M. *et al.* (2005) Structural characterization of components of protein assemblies by comparative modeling and electron cryo-microscopy. *J. Struct. Biol.*, **149**, 191–203.

Topf,M. *et al.* (2006) Refinement of protein structures by iterative comparative modeling and CryoEM density fitting. *J. Mol. Biol.*, **357**, 1655–1668.