BIOINFORMATICS APPLICATIONS NOTE Vol. 30 no. 22 2014, pages 3279–3280
doi:10.1093/bioinformatics/btu510

Structural bioinformatics

Advance Access publication July 26, 2014

Kotai Antibody Builder: automated high-resolution structural modeling of antibodies

Kazuo Yamashita[1], Kazuyoshi Ikeda[2], Karlou Amada[1], Shide Liang[1], Yuko Tsuchiya[2], Haruki Nakamura[3], Hiroki Shirai[4] and Daron M. Standley[1,*]

[1]WPI Immunology Frontier Research Center (IFReC), Osaka University, Suita, Osaka 565-0871, [2]National Institute of Biomedical Innovation, Ibaraki City, Osaka 567-0085, [3]Institute for Protein Research, Osaka University, Suita, Osaka 565-0871 and [4]Molecular Medicine Research Laboratories, Drug Discovery Research, Astellas Pharma Inc., Tsukuba, Ibaraki 305-8585, Japan

Associate Editor: Alfonso Valencia

ABSTRACT

Motivation: Kotai Antibody Builder is a Web service for tertiary structural modeling of antibody variable regions. It consists of three main steps: hybrid template selection by sequence alignment and canonical rules, 3D rendering of alignments and CDR-H3 loop modeling. For the last step, in addition to rule-based heuristics used to build the initial model, a refinement option is available that uses fragment assembly followed by knowledge-based scoring. Using targets from the Second Antibody Modeling Assessment, we demonstrate that Kotai Antibody Builder generates models with an overall accuracy equal to that of the best-performing semi-automated predictors using expert knowledge.

Availability and implementation: Kotai Antibody Builder is available at http://kotaiab.org

Contact: standley@ifrec.osaka-u.ac.jp

Received and revised on May 24, 2014; accepted on July 18, 2014

1 INTRODUCTION

Antibody variable regions constitute a unique protein module that has evolved to recognize virtually any biomolecular structure with high specificity and affinity. These properties have enabled the design of antibodies for use in the diagnosis and treatment of cancers and autoimmune and infectious diseases (Kuroda *et al*., 2012). In addition to their clinical value, antibodies are extremely important for routine assays used in basic research. Computational modeling of antibody structure is a crucial step in engineering new antibody molecules, but there are few tools available to the general public, and accurately modeling loops in complementary determining regions (CDRs) remains an open problem. The PIGS server (Marcatili *et al.,* 2008) was validated in the first blind Antibody Modeling Assessment (Almagro *et al*., 2011). However, prediction of the third heavy chain CDR (CDR-H3) remains difficult because of its structural diversity.

Recently, the Second Antibody Modeling Assessment (AMA-II) was held. AMA-II was divided into two stages: in stage 1, sequences were provided, and teams were assessed on the overall accuracy of their models. In stage 2, the crystal structures of the variable region lacking only CDR-H3 were provided, and groups were assessed on the accuracy of CDR-H3 loop prediction. In stage 1, the joint Osaka University Astellas (JOA) team achieved the lowest average root-mean square deviation (RMSD) for CDR-H3 (2.3 Å) and generated the most accurate models for 4 of 11 targets. In stage 2, the JOA team generated the most accurate models (with RMSDs of 1 Å or less) for 4 of 10 targets (Almagro *et al*., 2014). However, the method used by the JOA team required much manual intervention and expert knowledge. Kotai Antibody Builder represents a fully automated but simplified implementation of the pipeline used by the JOA team (Shirai *et al*., 2014).

*To whom correspondence should be addressed.

2 METHODS

Kotai Antibody Builder is composed of two main modules: MANGO and Spanner (Lis *et al*., 2011). The MANGO module selects template structures for the framework (i.e. non-CDR) and each CDR by a sequence-based database search and rule-based heuristics, while Spanner builds loops by fragment assembly. Because CDR-H3 loops are well known to be more difficult to model than those of other CDRs, Kotai Antibody Builder provides a refinement option that includes sampling by fragment assembly followed by side-chain modeling and scoring by an empirical scoring function.

2.1 Framework selection

The local structure of residues 7–10 in the heavy (H) chain (here denoted 'framework motif') is diverse and can be classified into five types. In the first step of Kotai Antibody Builder, the framework motif is predicted by a statistics-based classification. Next, sequence alignment is used to find framework templates for H and light (L) chains separately. Only templates that have the same framework motif with that predicted for the query are used. Here, the CDR regions are masked so that only the framework region is aligned and scored. The H and L results are merged and sorted by sequence identity (seqID). The Molprobity software (Chen *et al*., 2010) is used to assess if the percentage of backbone Ramachandran conformations inside the favored region is above a threshold (>85%). If there is no such model found, the selection criteria are relaxed and template models are selected by seqID, regardless of the predicted framework motif.

2.2 Non-H3 CDR selection

It is well known that each non-H3 CDR can be classified into one of several canonical clusters. We use the most recent definition of CDR clusters (North *et al*., 2011), along with position-specific scoring matrix

(PSSM)-based scoring, to predict the best non-H3 CDR cluster for the query (Shirai *et al*., 2014). For a given CDR, if the framework template and query are predicted to belong to the same cluster, the loop in the framework template is used; otherwise, the template with the highest seqID in the predicted cluster is used.

2.3 CDR-H3 selection

Because CDR-H3 is diverse in terms of length, sequence and structure, canonical rules have not been identified. Earlier, we developed H3 rules that partly classify CDR-H3 structures based on amino acid sequence (Kuroda *et al*., 2008; Shirai *et al*., 1996). Kotai Antibody Builder uses the most important of these rules (rule i), which predicts the structural class of the 'base' proximal to CDR-H3. For construction of the initial model, we use rule i if the length of loop is longer than five residues. In the optional refinement step, Spanner is used to generate 20 loop models followed by successive energy minimizations by the cosgene module in the myPresto package (Fukunishi *et al*., 2003) and OSCAR-leap (Liang *et al*., 2014). Here, we used a customized Spanner fragment database including only antibody structures for fast and specific CDR loop modeling. Finally, the single loop model was selected based on the OSCAR-leap score.

2.4 Model building

Selected loop models are grafted onto the framework template. If there are insertions/deletions in the template, Spanner is used to fix them. Side-chain modeling by OSCAR-star (Liang *et al*., 2011) followed by short energy minimization with positional restraints on backbone atoms is also carried out (Fukunishi *et al*., 2003). The initial calculation takes 5–10 min, whereas refinement requires an additional ~90 min.

2.5 Web server

Kotai Antibody Builder accepts amino acid sequences for H and L Fv regions. The resulting 3D model can be downloaded in PDB format and visualized by the JSmol viewer. The PDB IDs of templates as well as the canonical classifications of each loop are shown.

3 RESULTS

To assess the accuracy of Kotai Antibody Builder, we used targets 2–11 from AMA-II. In Figure 1, we show the resulting Cα RMSDs of rank-1 models submitted by the JOA team alongside rank-1 initial and refined models from the Web server. The overall RMSD was ~1 Å. The refinement option was much more successful in modeling CDR-H3 loops than the protocol used to generate initial models or by the PIGS server (Marcatili *et al.,* 2008). Surprisingly, the CDR-H3 accuracy for the refined loops (2.3 Å) was equal to that of the stage-1 JOA submitted models, the generation of which required careful manual inspection. There was a slight increase (~0.1 Å) in the RMSD of the non-H3 CDR loops when the refinement option was used because of the fact that the other loops were not held rigid during the minimizations; however, we found that this slight flexibility in the non-H3 CDRs was necessary for proper modeling of the H3 loops.

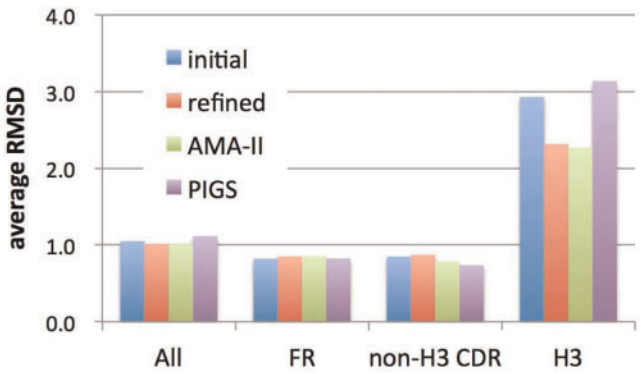

Fig. 1. Average Cα RMSDs of AMA-II targets 2–11. 'AMA-II' indicates rank-1 models submitted by the JOA team; 'initial' and 'refined' indicate the corresponding Web server options; 'All' and 'FR' indicate the entire Fv and framework regions for H and L chains combined

4 CONCLUSIONS

There are few fully automated antibody modeling pipelines available to the general public, and none that we are aware of that can reach this level of accuracy for CDR-H3 loops. Thus, Kotai Antibody Builder is expected to contribute uniquely to the field of antibody structural modeling and design.

ACKNOWLEDGEMENTS

The authors would like to thank N. Sakiyama, H. Nakagawa, E. Kanamori, K. Tsuchida, N. Tanigawa and S. Soga of Astellas Pharma for helpful discussions.

Funding: This work was supported by the Platform for Drug Discovery, Informatics and Structural Life Science, MEXT, Japanese Government.

Conflict of interest: none declared.

REFERENCES

Almagro,J.C. *et al.* (2011) Antibody modeling assessment. *Proteins*, **79**, 3050–3066.
Almagro,J.C. *et al.* (2014) Second antibody modeling assessment (AMA-II). *Proteins*, **82**, 1553–1562.
Chen,V.B. *et al.* (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.*, **D66**, 12–21.
Fukunishi,Y. *et al.* (2003) The filling potential method: a method for estimating the free energy surface for protein-ligand docking. *J. Phys. Chem. B.*, **107**, 13201–13210.
Kuroda,D. *et al.* (2008) Structual classification of CDR-H3 revisited: a lesson in antibody modeling. *Proteins*, **73**, 608–620.
Kuroda,D. *et al.* (2012) Computer-aided antibody design. *Protein Eng. Des. Sel.*, **25**, 507–521.
Liang,S. *et al.* (2011) Fast and accurate prediction of protein side-chain conformations. *Bioinformatics*, **27**, 2913–2914.
Liang,S. *et al.* (2014) LEAP: highly accurate prediction of protein loop conformations by integrating coarse-grained sampling and optimized energy scores with all-atom refinement of backbone and side chains. *J. Comput. Chem.*, **35**, 335–341.
Lis,M. *et al.* (2011) Bridging the gap between single-template and fragment based protein structure modeling using Spanner. *Immunome Res.*, **7**, 1–8.
Marcatili,P. *et al.* (2008) PIGS: automatic prediction of antibody structures. *Bioinformatics*, **24**, 1953–1954.
North,B. *et al.* (2011) A new clustering of antibody CDR loop conformations. *J. Mol. Biol.*, **406**, 228–256.
Shirai,H. *et al.* (1996) Structural classification of CDR-H3 in antibodies. *FEBS Lett.*, **399**, 1–8.
Shirai,H. *et al.* (2014) High-resolution modeling of antibody structures by a combination of bioinformatics, expert knowledge, and molecular simulations. *Proteins*, **82**, 1624–1635.