HLABAP: HLA Class I-Binding Antigenic Peptide Predictor

Ryoichi Kataoka¹, Shinji Amari¹, Takashi Ikegami¹, Noriaki Hirayama²*

¹ Life Science Department, MOLSIS Inc., 3-19-9, Hatchobori, Chuo-ku, Tokyo, 104-0032, Japan
² Institute of Advanced Biosciences, Tokai University, 143, Shimokasuya, Isehara, Kanagawa, 259-1193, Japan
*E-mail: hirayama@is.icc.u-tokai.ac.jp

(Received January 8, 2020; accepted February 7, 2020; published online March 2, 2020)

Abstract

HLA (Human Leucocyte Antigen) class I molecules present a variable but limited repertoire of antigenic peptides for T-cell recognition. Identification of specific antigenic peptides is essential for the development of immunotherapy. High polymorphism of HLA genes and a large number of possible peptides to be evaluated, however, have made the identification by experiments costly and time-consuming. Computational methods have been proposed to address this problem. In cases where plenty number of binding affinity data of peptides are available, various QSAR and machine learning approaches efficiently evaluate the affinity of test peptides, while in the cases where just a little data are available, structure-based approaches like elaborate docking have been proposed. We have developed a software named HLABAP that is designed to predict the binding affinities for a set of peptides against a particular HLA class I allele. By the combination of homology modeling for posing instead of docking and geometry optimization of the complex structures between the HLA molecule and peptides, HLABAP well predicts the binding affinities for the peptides. The results have shown that HLABAP should be applicable to identify possible antigenic peptides against a particular allele of HLA class I prior to the experiments far efficiently than the ordinary docking methods.

Key Words: HLA, Antigenic Peptides, Homology Modeling, HLA-A*02:01, MOE

Area of Interest: Molecular recognition and molecular modeling

1. Introduction

Cellular immune mechanisms detect and destroy cancerous and infected cells via the HLA class I molecules that present antigenic peptides of intracellular origin on the surface of nucleated cells. The identification of tumor-specific epitopes is an essential step in the development of immunotherapy for malignant tumors. Antigenic peptides presented by HLA class I are mainly nonamers. High polymorphism of HLA genes and a large number of possible peptides to be evaluated, however, makes experimentally testing costly and time-consuming. Therefore, computational methods have been proposed to address this problem [1]. As most of the currently available methods are allele-specific and dependent on the number of experimentally determined binding affinity data,
they cannot be applied to the cases without sufficient training samples. In this study, we propose a pan-specific prediction method based on the optimized 3D structures of the complex between the HLA class I molecule and a set of potential peptides or proteins. Since the binding affinity can be evaluated by the estimated physicochemical properties such as binding free energy, the method could be generally applicable to any HLA class I molecules without preliminarily obtained experimental data.

2. Methods

The application HLABAP is coded using built-in scientific vector language in the integrated computer-aided molecular design platform MOE [2]. HLABAP constructs the optimal 3D structures of the complexes between a given HLA allele and peptides using the same MOE Homology Model method as HLA-Modeler [3], a homology modeling application specifically oriented to HLA molecules. The method is based on segment matching [4] and the rotamer assembly of side chains. Execution steps of HLABAP are as follows:

(1) Preparing the 3D structure of the template HLA molecule
The most appropriate 3D model of the complex between the template HLA molecule and an antigenic peptide is obtained from the Protein Data Bank [5] or constructed using HLA-Modeler.

(2) Assorting a set of amino acid sequences of antigenic peptides or proteins to be evaluated

(3) Aligning antigenic peptides with the template peptide with tethering N- and C-terminal residues and building homology models of the peptides on the HLA molecule
When a protein sequence is given, it is threaded with the nine residues-long windows in order to predict antigenic moiety.

(4) Optimization of the 3D structures of the HLA-peptide complexes
The HLA-peptide complex structures are optimized with Amber10:EHT force field and Generalized Born solvation model in MOE.

(5) Calculating affinity scores of the optimized structures in order to identify the peptides with high affinity to the target HLA molecule
Scoring functions of Contact Energy [6], Affinity dG, London dG, and GBVI/WSA dG [7] available in MOE are calculated. Contact Energy estimates the transfer free energy based on the area of the water-accessible surface. Affinity dG is an empirical scoring function evaluating the directional hydrogen-bonding, the directional hydrophobic interaction, and the entropy (ligand rotatable bonds immobilized in binding). London dG estimates the free energy of binding summing the rotational and translational entropy, the energy loss of flexibility of the ligand, the geometric imperfections of hydrogen bonds, the geometric imperfections of metal ligations, and the desolvation energy. GBVI/WSA dG is a forcefield-based scoring function trained using the MMFF94x and AMBER99 forcefield on the 99 protein-ligand complexes of the solvated interaction energy training set. In addition to each single scoring function, consensus scoring functions were considered.

A substantial number of HLA-A*02:01-restricted cytotoxic T lymphocyte epitopes closely related to tumors have been identified so far [8]. In order to illustrate the performance of HLABAP, the experimentally determined pIC$_{50}$ values and the predicted binding affinities of the peptides against HLA-A*02:01 were compared in this study. The 119 crystal structures of complexes between HLA-A*02:01 and antigenic peptides obtained from the Protein Data Bank were used. The experimental affinity data (IC$_{50}$) of 3,995 peptides were obtained from the Immune Epitope Database (IEDB) [9].

3. Results and discussion
Modeling the appropriate structure of the complex between the HLA molecule and peptides is the crucial stage in predicting the pIC$_{50}$ values. For the 119 crystal structures, the root-mean-square deviations (RMSDs) of the backbone atoms of the peptides between the crystal structures and predicted ones by HLABAP were calculated and the results are summarized in Figure 1. Out of nine amino acid residues, the C- and N-terminal residues are important in anchoring the peptide to the antigenic-peptide binding groove of the target HLA molecule. The RMSD values of these residues are small enough showing that the structures of anchoring residues are adequately predicted by HLABAP. These residues are expected to greatly contribute to the pIC$_{50}$ values. Since the P3 to P7 residues are exposed on the surface of the HLA molecule, these residues take relatively flexible conformations if the corresponding T-cell receptors are not bound. The results indicate that HLABAP can predict the 3D structures of the antigenic peptides decently accurate for predicting the pIC$_{50}$ value.

**Figure 1.** Backbone RMSDs of the peptide between the crystal structures and the predicted ones by HLABAP for the 119 crystal structures of the complexes between HLA-A*02:01 and antigenic peptides. The amino acid residues of the nonapeptides are numbered P1 to P9.

The correlations between the pIC$_{50}$ values and the single scoring functions are given in Figure S1 and Table 1. Contact Energy is the most correlated scoring function among the calculated ones. The calculated area under the ROC curve (AUC) is 0.748, suggesting that the pIC$_{50}$ value can be adequately predictable for any HLA class I allele by HLABAP. Given that that pIC$_{50}$ values are assay specific and therefore not compatible between different assays done in different laboratories [10], the prediction accuracy by use of Contact Energy should be acceptable.

**Table 1.** Correlations between the experimental pIC$_{50}$ values and single scoring functions

|                  | Contact Energy | Affinity dG | GBVI/WSA dG | London dG |
|------------------|----------------|-------------|-------------|-----------|
| $r^2$            | 0.301          | 0.105       | 0.067       | 0.007     |
| AUC              | 0.748          | 0.669       | 0.631       | 0.569     |

As each scoring function has its strengths and its weaknesses, the discriminatory ability of a single scoring alone is not necessarily good enough in certain cases. To overcome the limitation of a single scoring function, consensus scoring functions have been applied to improve discriminating performance [11]. In this study, we assessed the discrimination ability of three consensus scores obtained by combining Contact Energy and each one of the remaining three scoring functions. The optimized coefficients attributed to individual scores are given in Equation S1. The prediction accuracy is summarized in Figure S2 and Table S1. The consensus scores of Contact Energy + London dG and Contact Energy + GBVI/WSA dG appreciably improve predictive accuracy. The results indicate that London dG and GBVI/WSA dG could compensate for the energetic contribution...
of hydrogen bonds that is less weighted in Contact Energy.

Taken collectively, the present study suggests that HLABAP should provide a reasonably accurate prediction of the binding affinities of a set of peptides against a particular HLA class I allele, and should thus help us to identify the antigenic peptides prior to the time-consuming experiments.

4. Acknowledgments

This work was partly supported by Grant-in-Aid for Scientific Research on Innovative Areas (22133012) from the Ministry of Education, Culture, Sports, Science, and Technology (MEXT) for Noriaki Hirayama.

5. Supplements

Figure S1, Figure S2, Equation S1 and Table S1 is available at: https://doi.org/10.1273/cbij.20.1

References

[1] Gfeller, D.; Bassani-Sternberg, M. Predicting Antigen Presentation—What Could We Learn From a Million Peptides? *Front. Immunol.* **2018**, *9*, 1–17. doi: 10.3389/fimmu.2018.01716
[2] Molecular Operating Environment (MOE), 2019.01; Chemical Computing Group ULC, 1010 Sherbrooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7, **2019**.
[3] Amari, S.; Kataoka, R.; Ikegami, T.; Hirayama, N. HLA-Modeler: Automated Homology Modeling of Human Leukocyte Antigens. *Int. J. Med. Chem.* **2013**, vol. 2013, Article ID 690513, 6 pages. doi: 10.1155/2013/690513
[4] Levitt, M. Accurate modeling of protein conformation by automatic segment matching, *J. Mol. Biol.* **1992**, 226, 507–533. doi: 10.1016/0022-2836(92)90964-1
[5] Bernstein, F. C.; Koetzle, T. F.; Williams, G. J.; Meyer, E. F. Jr.; Brice, M. D.; *et al.* The protein data bank: A computer-based archival file for macromolecular structures. *J. Mol. Biol.* **1977**, **112**, 535–542. doi: 10.1016/s0022-2836(77)80200-3
[6] Zhang, C.; Vasmatzis, G.; Cornette, J. L.; Delisi, C. Determination of Atomic Desolvation Energies From the Structures of Crystallized Proteins. *J. Mol. Biol.* **1997**, **267**, 707–726. doi: 10.1006/jmbi.1996.0859
[7] Kalinowsky, L.; Weber, J.; Balasupramaniam, S.; Baumann, K.; Proschak, E. A Diverse Benchmark Based on 3D Matched Molecular Pairs for Validating Scoring Functions. *ACS Omega* **2018**, *3*, 5704–5714. doi: 10.1021/acsomega.7b01194
[8] Yoshimura, M.; Tada, Y.; Ofuzi, K.; Yamamoto, M.; Nakatsura, T. Identification of a novel HLA-A02:01-restricted cytotoxic T lymphocyte epitope derived from the EML4-ALK fusion gene. *Oncol. Rep.* **2014**, *32*(1), 33–39. doi: 10.3892/or.2014.3198
[9] The Immune Epitope Database (IEDB), **2019**, http://www.iedb.org/
[10] Kalliokoski, T.; Kramer, C.; Vulpetti, A.; Gedeck, P. Comparability of Mixed IC50 Data – A Statistical Analysis. *PLoS ONE* **2013**, **8**(4): e61007. doi: 10.1371/journal.pone.0061007
[11] Vieira, T. F.; Magalhaes, R. P.; Sousa, S. F. Tailoring specialized scoring functions for more efficient virtual screening. *Front. Drug Chem.Clin. Res.* **2019**, **2**, 1–4. doi: 10.15761/FDCCR.1000118