An Interaction-based Approach for Affinity Prediction between Antigen Peptide and Human Leukocyte Antigen Using COMBINE Analysis

Shinya Nakamura*, Rie Ohmura, Isao Nakanishi

Department of Pharmaceutical Sciences, Kindai University, 3-4-1 Kowakae, Higashiosaka, Osaka 577-8502, Japan

**E-mail: nakas@phar.kindai.ac.jp*

(Received August 19; accepted August 30; published online September 6, 2017)

Abstract

In peptide vaccine therapy, a peptide with high affinity for human leukocyte antigen (HLA), is important to stimulate the immune system to kill cancer cells. Several methods to predict HLA–peptide binding have been reported, but most of them rely on informatics to analyze the amino acid sequence of the peptide. Although intermolecular-interaction-based analysis is expected to improve prediction accuracy, such a method generally involves a high computational cost. Therefore, comparative binding energy (COMBINE) analysis, a 3D-quantitative structure–activity relationship method, combined with a rapidly implemented protein modeling method, was applied to solve this problem. The new method enabled quick evaluation of peptide affinity predictions with accuracy beyond a statistical method. In addition, several amino acid residues of HLA, which are known to be important for peptide binding, could be identified.

Key Words: Peptide vaccine therapy, Human leukocyte antigen (HLA), Comparative binding energy (COMBINE) analysis, Quantitative structure–activity relationship

Area of Interest: *In silico* drug discovery

1. Introduction

A considerable number of medicines have been developed to improve cancer treatment [1]. Cancer immunotherapy [2] or immunooncology is used to treat patients with cancer, an approach involving or using components of the patient's own immune system. In peptide vaccine therapy [3], a peptide fragment that is highly expressed specifically in cancer cells is administered to the patient as an antigen to activate cytotoxic T-lymphocytes (CTLs). The T-cell receptor (TCR) of CTLs recognizes the peptide fragments, which become markers of cancer cells; CTLs are thus activated and attack cancer cells through cytotoxicity. At this time, it is important to note that since the TCR recognizes the complex formed between the peptide fragment and human leukocyte antigen (HLA), instead of the peptide fragment alone (Fig. 1(a)), both the affinity between HLA and the peptide fragment and that between TCR and the HLA-peptide fragment complex are important. However, predicting the TCR-based complex structure and affinity is a very complex task because TCRs have very diverse sequences to respond to a wide variety of antigens [4].

Several tools have been reported for the prediction of antigen-peptide recognition. They are based on various statistic or bioinformatics approaches, including the hidden Markov model method [5], the machine learning method [6], and the artificial neural network method [7]. For example, the web tool called BioInformatics and Molecular Analysis Section (BIMAS) [8] predicts peptide activity based on the half-life ($T_{1/2}$) of β 2-microglobulin (β 2M) separation from a complex comprising a peptide fragment and HLA (Fig. 1(a)). This half-life is reported to be an indicator of the stability of the HLA-peptide complex because peptide-induced stabilization of HLA also indirectly stabilizes the binding between HLA and β 2M [9]. The database called SYFPEITHI [10] predicts peptide activity based on a score that indicates the presence or absence of anchor residues, which are important for the binding affinity between the peptide fragment and HLA. MHCPre [11], a web server for quantitative prediction of peptide-MHC binding, uses a parameter determined simply from the species and sequential position of the residue. As these statistical tools are based on the search of motifs that can be recognized as antigens or epitopes, the interaction between the peptide and HLA at the molecular or atomic level is not considered.

On the other hand, several structure-based approaches, which construct models of the HLA-peptide complex using docking simulations and estimate binding affinities through theoretical energy calculation, have been reported [12][13]. Implementing these methods is time-consuming and involves a huge computational cost. In addition, serious problems in accurate modeling of peptide-protein interactions have been pointed out. In fact, since current docking methodologies have been optimized to estimate the interaction between small molecules and proteins, many of such methodologies cannot be applied to the estimation of the interaction between proteins and immunogenic peptides, which are larger and more flexible than "drug-like" compounds [13]. Therefore, the introduction of smart methods characterized by both accuracy and low computational cost is desirable.

In this study, to accelerate the affinity calculation, an interaction-based method that uses comparative binding energy (COMBINE) analysis [14] was adopted. COMBINE analysis is a three-dimensional quantitative structure-activity relationship (QSAR) method. Such methods are generally used for estimating the binding affinity of not only protein-synthetic ligand complexes but also of protein-peptidyl ligand complexes. In COMBINE analysis, the interaction energy between ligand and protein is broken down into its components, in particular into those relating to each amino acid residue unit. Once an appropriate prediction model is established, the computational cost of predicting the affinity of new ligands for a particular protein is extremely low.

On the other hand, for the efficient modeling of peptide–HLA complex, in the approach adopted in this study, a rotamer conformational search technique was applied, which is much faster to implement than docking methods. Binkowski *et al.* examined the root-mean-square deviation (RMSD) values of 50 peptides for which the structures of the complexes with HLA-A2 (A*0201) were experimentally determined and found that the average RMSD is less than 1.0 Å [13]. This evidence suggested that an approximate modeling method might suffice to obtain acceptably accurate affinity predictions in the present context.

In this study, the accuracy of the interaction-based affinity prediction using the COMBINE method was compared to that of a statistical method, BIMAS. As a benchmark test dataset to validate the new method, 81 peptide sequences and the $T_{1/2}$ values used for BIMAS parameterization [8] were utilized for the analysis.

Figure 1. Interaction between HLA and the peptide fragment

(a) Schematic representation of a complex comprising a peptide fragment and an MHC class I molecule on an antigen presenting cell and TCR on a cytotoxic T-lymphocyte. (b) Schematic model and molecular structure example for peptide–HLA complex to calculate the interaction energy.

2. Materials and Methods

As a modeling template, the structure of the complex that forms between human class I major histocompatibility complex (MHC) HLA-A2 (A*0201) and an epitope HuD peptide fragment (LGYGFVNYI) was retrieved from the Protein Data Bank (PDB; PDB-ID: 3PWL [15]). Since the approach adopted herein involves a higher computational cost than do bioinformatics tools, ways to reduce the cost of implementing such an approach preparation of an appropriate HLA–peptide complex model should be considered. Class I MHC comprises four domains: HLA-α1 and HLA-α2 recognize antigens, HLA-α3 anchors the structure to the surface of the cell membrane, and β2M associates with the HLA-α chain (see Fig. 1(a)). In a preliminary study, interaction energies between the LGYGFVNYI peptide and each of the four domains were calculated based on the PDB-retrieved structure. The calculated interaction energies of the peptide with the HLA-α3 and β2M domains were approximately zero (i.e., negligibly weak, compared to those with the HLA-α1 and HLA-α2 domains). As a consequence, the complex formed between the peptide and the HLA-α1-α2 domains was assumed to be the peptide–HLA complex in all subsequent simulations (see Fig. 1(b)).

As mentioned in the Introduction, several computational methods exist for modeling a complex structure. For efficient modeling, a rotamer conformation search of peptide side chains was performed using the Rotamer Explorer module implemented in MOE [16], expecting tolerable structural accuracy for affinity prediction. In this scheme, the hydrogen bond network involving peptide main-chain atoms was assumed to be the same as that of a template peptide–HLA complex, regardless of side chains.

In this study, a total of 81 nonameric peptides used in the parametrization of BIMAS [8] were examined to compare the performance of the interaction-based approach with that of a residue-based statistical approach, BIMAS. After constructing the 81 complex structures, assignment of the AMBER99 force-field parameters [17], geometry optimization, and estimation of the binding affinity of the peptide fragment were subsequently carried out. A three-step energy minimization procedure was performed for geometry optimization: the positions of hydrogen atoms, side chain atoms, and, finally, all atoms, with a position restraint of $100 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$ for main-chain atoms, were gradually optimized. For each step, the optimization was performed by implementing the steepest descent method, conjugate gradient method, and truncated Newton method for 100, 100, and 500 steps, respectively, or until the root-mean-square gradient was below $0.01 \text{ kcal mol}^{-1} \text{ \AA}^{-1}$. A cutoff distance of $10 \text{ \AA}$ was set for nonbonding interactions, and the solvation effect was evaluated by the Generalized Born/Volume Integral (GB/VI) method [18].

In COMBINE analysis [14], the interaction energy between a peptide and the HLA protein is broken down into the contributions of the individual amino acid residues, and, thus, the interaction energy of the whole peptide as a ligand to each residue of HLA was calculated. The PLS regression [19] models for predicting the activities of peptides were obtained by COMBINE analysis using the COMBINE software [14]. As a measure of the activity of a peptide was utilized $\log T_{1/2}$, the logarithm of $\beta 2\text{M}$'s dissociation half-time ($T_{1/2}$), which is an indicator of the stability of the HLA–peptide complex. First, expected predictabilities of all 81 peptides were estimated as external validations using a leave-several-out validation scheme: a quarter of the 81 peptides were randomly chosen as an external test set to predict activity values; a prediction model was then obtained from the remaining peptides used as a training set. This process was repeated four times. The optimal number of latent variables (LVs) in PLS regression was determined so that the regression model have the highest Q^2 value obtained by internal leave-one-out cross validation for training set. Subsequently, a prediction model with all 81 peptides was constructed to compare its accuracy with that based on BIMAS analysis. In addition, permutation analysis by scrambling test of the randomly shuffled activities was performed 100 times to check the case of by-chance correlation model.

3. Results and Discussions

After constructing 81 peptide-based complexes by rotamer search, the relationship between their experimental activities (indicated by the value of $\log T_{1/2}$ [9]) and those calculated by COMBINE analysis was investigated. The predictabilities estimated by leave-several-out validation are shown in Table 1 (four runs and their average are named LSO Runs 1–4 and LSO Ave., respectively). In general, the threshold Q^2 value for a good prediction model is >0.5 [20]. Although the average Q^2 value of the four runs was 0.48, the Q^2 values in LSO Run 2 and Run 3 exceeded the threshold. Thus, a good prediction was expected, and, at the very least, the accuracy achieved through this approach is expected to be higher than that achieved with MHCPre, which is a QSAR method (Q^2 : 0.38) that does not take into consideration molecular interactions [11].

Table 1. Prediction model construction in leave-several-out validation set and all data^a Number of peptides used in the training set to construct the regression model.^b Number of peptides used in the test set to predict peptide activity using the regression model.^c Optimal number of latent variables of PLS regression with the best Q^2 values.^d Q^2 value as an expected predictability was calculated using internal leave-one-out cross validation.^e Root mean square of prediction error (RMSE) for predicting internal leave-one-out cross validation.^f RMSE for predicting the external test set.

	#Train ^a	#Test ^b	LV ^c	Training set			Test set	
				R^2	Q^2 ^d	RMSE _{CV} ^e	R^2	RMSE _{pred} ^f
LSO Run 1	61	20	6	0.72	0.38	0.69	0.78	0.37
LSO Run 2	61	20	6	0.77	0.51	0.56	0.32	0.63
LSO Run 3	61	20	6	0.80	0.58	0.59	0.35	0.64
LSO Run 4	60	21	6	0.75	0.44	0.58	0.49	0.54
LSO Ave.	—	—	—	0.76	0.48	0.56	0.49	0.55
Final	81	0	6	0.79	0.52	0.54	—	—

Figure 2. Permutation analysis of prediction model (Final)

R^2 (horizontal axis) and Q^2 (vertical axis) values for correct prediction model and 100 scrambled prediction models were shown in closed square and open triangles, respectively.

Finally, a regression model was constructed using all 81 peptides (named Final in Table 1). As the number of peptides in the training set increased, the Q^2 value exceeded the threshold of 0.5, a better prediction than the QSAR method without considering molecular interactions [11]. Based on the result of permutation analysis (Fig. 2), there was no evidence of chance correlation, and a robust prediction model had been constructed.

The prediction accuracies of data obtained following implementation of the BIMAS and COMBINE analyses were compared (Fig. 3 and Table 2): Two peptides are obvious outliers in the BIMAS approach, whereas the activities of all peptides could be accurately predicted in the COMBINE-based analysis. BIMAS underestimated the activity of peptide GKLGFVFTL (experimental $\log T_{1/2}$: 3.176; BIMAS: 0.745; COMBINE: 2.778) and overestimated that of peptide GLGGGGFGV (experimental $\log T_{1/2}$: 0.903; BIMAS: 2.606; COMBINE: 1.651). For high activity (large $\log T_{1/2}$ value in this study), the identity of the second amino acid residue of the HLA-A2 epitope peptide has been reported to have a Leu or an Ile residue [21-23]. Peptide GKLGFVFTL

(see above) displays high activity even though its second residue is a Lys (K), a topology that does not meet the preference criteria mentioned above. Conversely, GLGGGGFGV has low activity even though the second residue is a Leu (L), which in principle is a favorable situation. There was only one peptide whose second residue was a Lys among the 81 peptides tested. A residue-based statistical method might have largely overestimated the effect of this rare residue, whereas COMBINE analysis could well reproduce the experimental value.

COMBINE analysis also provides information on the relative importance of each amino acid residue of the host molecule with respect to activity [14]. PLS pseudo coefficients (simply called “weights”) of each residue in this COMBINE prediction model (Final) are reported in Fig. 4. The residues with strong weight peaks, which have large absolute values, are important for peptide activity differences in this data set. Several reports exist on the biological or medical effects of HLA amino acid residue mutations [24-26]. The importance of most residues, as determined by COMBINE analysis, is consistent with experimental observation; for example, Tyr99 and Glu63, characterized by very strong peak weights, were known as important residues [24]. However, physiologically important residues, such as Leu156, Tyr116, and Phe9 [26], were also detected by COMBINE analysis with weak peak weights. In particular, the weight of Phe9, which constitutes the bottom surface near the center of the peptide-binding pocket, is considerably weak. This residue interacts with the bulky side chain of the second residue of the peptide, such as the Leu residue mentioned above. Since both highly active and weakly active peptides interacted with Phe9, this residue would have been interpreted as having little influence on activity. In this regard, the weights reflect the influence that the relevant residues have on activity. Notably, Tyr7 had the most significant coefficient in COMBINE analysis. Tyr7 is a conserved residue in other HLA families [27][28], and its interactions are reportedly maintained in molecular dynamics simulations [29]. Since, to the best of our knowledge, no physiological experimental evidence has been reported on the importance of this residue, we believe that this result might be worth pursuing by researchers.

Figure 3. Correlation between the calculated and experimental $\log T_{1/2}$ values

Results of COMBINE and BIMAS analyses are shown as closed squares and open circles, respectively.

Figure 4. PLS pseudocoefficient/weight (vertical axis) of each residue in COMBINE analysis

Residue number is shown on the horizontal axis. Weak weights were omitted by showing them as having a zero value for clarity. Residues with strong weights are labeled and those referred to in the text were labeled with parentheses.

4. Conclusion

In peptide vaccine therapy, a peptide with high affinity for HLA is important to stimulate the immune system to kill cancer cells. Several HLA–peptide binding prediction methods have been reported previously, but most of these methods rely on informatics to analyze the amino acid sequence of the peptide. A more precise analysis at the molecular level would be expected to improve the accuracy of the binding prediction. In this study, COMBINE analysis was performed as a method to obtain a quantitative structure–activity relationship. HLA–peptide complex models were constructed assuming that the hydrogen bond network of the main-chain atoms in the template structure was maintained, regardless of the specific side chains.

With comparable data set numbers, interaction-based COMBINE analysis was able to produce better predictive models than statistical methods and QSAR-based methods that did not take into consideration molecular interactions. The residue-based statistical prediction method is difficult to apply to the case of residues for which no experimental evidence exists; however, the molecular-interaction-based method would achieve robust predictions even in such a case, as the level of interaction can be calculated solely from their structure.

Table 2. The experimental and calculated log $T_{1/2}$ values of peptides in this study

sequence	Exp.	BIMAS	COMBINE	sequence	Exp.	BIMAS	COMBINE
GLLGFVFTL	4.176	3.603	3.392	GILGEVFTL	2.342	2.161	2.422
LMFGYPVYV	3.863	3.240	3.375	KLGEFYNQM	2.342	2.609	2.510
LLFGYPVYV	3.806	3.381	3.601	GILGFVFVL	2.322	2.354	2.882
CLFGYPVYV	3.663	3.151	3.277	FIFPNYTIV	2.301	2.000	2.865
KMFGYPVYV	3.643	3.554	3.824	GIVGFVFTL	2.255	2.173	1.787
KILGFVFTL	3.398	3.285	3.334	GLFGGGGGV	2.255	1.803	2.185
KILGKVFTL	3.322	2.705	2.992	GLYGGGGGV	2.146	1.461	1.682
GLFGGGGGV	3.301	3.174	2.599	NIVAHTFKV	2.146	1.744	2.164
LQFGYPVYV	3.279	2.387	2.929	GLFGGVGGV	2.079	1.886	2.137
AILGFVFTL	3.255	2.741	2.982	GLFGGGGGV	2.041	1.524	2.070
GKLGFVFTL	3.176	0.745	2.778	QMLLAIAL	2.041	2.118	1.831
GILGFVFKL	3.079	2.741	2.974	GLFKGGGGV	2.017	1.524	2.212
GILPFVFTL	3.041	2.741	2.865	GLFKGVGGV	2.000	1.886	2.142
GILGFVFAL	3.000	2.741	2.782	TLNAWVKV	2.000	1.817	2.234
GILGFVFTL	3.000	2.741	2.947	GILGFVATL	1.987	1.978	2.605
GILGFVFEL	2.940	2.741	3.285	GLFGGVGKV	1.982	1.886	2.148
GLFGGGGV	2.919	2.411	2.762	WLYRETCNL	1.959	2.498	2.515
GLFGGGAGV	2.886	2.411	2.168	GLLGGGVGV	1.954	2.411	2.599
GILGFEFTL	2.863	2.379	2.711	GALGFVFTL	1.949	1.745	2.085
KLFGGVGGV	2.863	2.430	2.926	GMFGGVGGV	1.924	1.745	1.213
GILAFVFTL	2.785	2.741	2.973	RLVNGSLAL	1.908	1.330	2.181
ALFAAAAV	2.756	2.411	2.377	GILGFKFTL	1.851	2.379	2.537
YLFKRMIDL	2.756	2.922	2.633	GILGFVKTL	1.833	1.978	1.440
GILGFVFTA	2.740	2.108	2.720	ALFFFDIDL	1.826	2.478	1.789
GLFGGGGV	2.724	2.101	2.868	GILGFVETL	1.813	1.668	1.997
NLVPMTATV	2.681	2.204	2.894	AILLGVFML	1.732	2.161	1.796
KLFGGGGGV	2.672	2.069	2.614	LNFGYPVYV	1.613	1.196	2.068
GILGKVFTL	2.643	2.161	2.169	GAIGFVFTL	1.580	1.177	1.587
QMWQARLTV	2.623	2.620	2.699	GIEGFVFTL	1.380	0.979	1.623
GQLGFVFTL	2.613	2.609	2.649	GLFGGGGGL	1.380	1.012	1.234
KALGFVFTL	2.613	2.290	2.761	GMFGGGGGV	1.342	1.383	1.761
SLYNTVATL	2.568	2.197	1.913	GQFGGVGGV	1.041	0.892	1.091
LLWKGEAV	2.556	2.106	2.316	ILESIFAAV	0.908	1.158	1.471
KLGEFFNQM	2.544	2.712	2.483	GIFGGGVGGV	0.903	1.024	1.054
ILASLFAAV	2.542	2.352	2.244	GLGGGGFGV	0.903	2.606	1.651
RLLQTGIHV	2.477	2.411	2.560	AIYKRWIIL	0.845	0.816	1.114
GELGFVFTL	2.447	1.892	2.585	ATVELLSFL	0.778	1.187	0.927
GILKFVFTL	2.398	2.741	2.736	GLDKGGGGV	0.778	0.558	1.204
GIAGFVFTL	2.380	2.173	1.932	GVFGGVGGV	0.778	0.828	0.841
GILFGAFTL	2.342	1.800	2.846	IISLWDQSL	0.778	0.698	0.183
GILGAVFTL	2.342	2.161	2.389				

However, the accuracy of the proposed method is not high enough. It is expected that prediction accuracy improve as the number of data increases. Therefore, more data on the affinity between HLA and peptides are required. The Immune Epitope Database (IEDB) [30] contains many assay results on various antigens and antibodies. Herein, the activities of most peptides in peptide vaccine therapy are rated on the basis of the value of the half-maximal effective concentration (EC_{50}). This value is affected not only by the interaction between HLA and peptide but also by the interaction between the TCR in CTL and the HLA-peptide complex. Since the ligand recognition region of the TCR is highly diverse, the number of combinations of peptide sequences and TCR sequences would be too high to handle by a structure-based approach. In a preliminary crude study, approximately 3000 HLA-peptide complexes with known EC_{50} values have been modeled and their interaction energy calculated, obtaining no correlation between interactions and EC_{50} values. This evidence suggests that utilization of EC_{50} or IC_{50} values as parameters reflecting the affinity between TCR and peptides is not adequate. As there are approximately 500 peptides with dissociation constants, K_d , that are strongly correlated with binding stability and interaction, we anticipate that a more accurate prediction model will be constructed using this data set.

Acknowledgement

This study was supported by the MEXT-Supported Program for the Strategic Research Foundation at Private Universities, 2014–2018(S1411037).

References

- [1] FDA Approved Drugs for Oncology, (accessed Aug 19, 2017). <https://www.centerwatch.com/drug-information/fda-approved-drugs/therapeutic-area/12/oncology>
- [2] Farkona, S.; Diamandis, E. P.; Blasutig, I. M. Cancer immunotherapy: the beginning of the end of cancer? *BMC Med.* **2016**, *14*, 73.
- [3] Tsukahara, T.; Hirohashi, Y.; Kanaseki, T.; Nakatsugawa, M.; Kubo, T. *et al.* Peptide vaccination therapy: Towards the next generation. *Pathol. Int.* **2016**, *66*, 547-553.
- [4] Laydon, D. J.; Bangham, C. R.; Asquith, B. Estimating T-cell repertoire diversity: limitations of classical estimators and a new approach. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **2015**, *370*, 20140291.
- [5] Zhang, C.; Bickis, M. G.; Wu, F. X.; Kusalik, A. J. Optimally-connected hidden markov models for predicting MHC-binding peptides. *J. Bioinform. Comput. Biol.* **2006**, *4*, 959-980.
- [6] Luo, H.; Ye, H.; Ng, H. W.; Shi, L.; Tong, W. *et al.* Machine Learning Methods for Predicting HLA-Peptide Binding Activity. *Bioinform. Biol. Insights* **2015**, *Suppl.3*, 21-29.
- [7] Andreatta, M.; Nielsen, M. Gapped sequence alignment using artificial neural networks: application to the MHC class I system. *Bioinformatics* **2016**, *32*, 511-517.
- [8] Parker, K. C.; Bednarek, M. A.; Coligan, J. E. Scheme for ranking potential HLA-A2 binding peptides based on independent binding of individual peptide side-chains. *J. Immunol.* **1994**, *152*, 163-175.
- [9] Parker, K. C.; DiBrino, M.; Hull, L.; Coligan, J. E. The beta 2-microglobulin dissociation rate is an accurate measure of the stability of MHC class I heterotrimers and depends on which peptide is bound. *J. Immunol.* **1992**, *149*, 1896-1904.
- [10] Rammensee, H. G.; Bachmann, J.; Emmerich, N. P. N.; Bachor, O. A.; Stevanovic, S. SYFPEITHI: database for MHC ligands and peptide motifs. *Immunogenetics* **1999**, *50*, 213-219.
- [11] Guan, P.; Doytchinova, I. A.; Zygouri, C.; Flower, D. R. MHCpred: A server for quantitative

- prediction of peptide-MHC binding. *Nucleic Acids Res.* **2003**, *31*, 3621-3624.
- [12] Bordner, A. J.; Abagyan, R. *Ab initio* prediction of peptide-MHC binding geometry for diverse class I MHC allotypes. *Proteins* **2006**, *63*, 512-526.
- [13] Binkowski, T. A.; Marino, S. R.; Joachimiak, A. Predicting HLA class I non-permissive amino acid residues substitutions. *PLoS One* **2012**, *7*, e41710.
- [14] Ortiz, A. R.; Pisabarro, M. T.; Gago, F.; Wade, R. C. Prediction of Drug Binding Affinities by Comparative Binding Energy Analysis. *J. Med. Chem.* **1995**, *38*, 2681-2691.
- [15] Borbulevych, O. Y.; Piepenbrink, K. H.; Baker, B. M. Conformational melding permits a conserved binding geometry in TCR recognition of foreign and self molecular mimics. *J. Immunol.* **2011**, *186*, 2950-2958.
- [16] MOE Ver. 2013, *Chemical Computing Group Inc.*, Montreal, Canada.
- [17] Wang, J.; Cieplak, P.; Kollman, P. A. How Well Does a Restrained Electrostatic Potential (RESP) Model Perform in Calculating Conformational Energies of Organic and Biological Molecules? *J. Comput. Chem.* **2000**, *21*, 1049-1074.
- [18] Labute, P. The generalized Born/volume integral implicit solvent model: estimation of the free energy of hydration using London dispersion instead of atomic surface area. *J. Comput. Chem.* **2008**, *29*, 1693-1698.
- [19] Wold, S.; Sjöström, M.; Eriksson, L. PLS-regression: a basic tool of chemometrics. *Chemometr. Intel. Lab. Syst.* **2001**, *58*, 109-130.
- [20] Golbraikh, A.; Tropsha, A. Beware of q^2 ! *J. Mol. Graph. Model.* **2002**, *20*, 269-276.
- [21] Falk, K.; Röttschke, O.; Stevanovic, S.; Jung, G.; Rammensee, H. -G Allele-specific motifs revealed by sequencing of self-peptides eluted from MHC molecules. *Nature* **1991**, *351*, 290-296.
- [22] Kubo, R. T.; Sette, A.; Grey, H. M.; Appella, E.; Sakaguchi, K. *et al.* Definition of specific peptide motifs for four major HLA-A alleles. *J. Immunol.* **1994**, *152*, 3913-3924.
- [23] Davenport, M. P.; Smith, K. J.; Barouch, D.; Reid, S. W.; Bodnar, W. M. *et al.* HLA Class I Binding Motifs Derived from Random Peptide Libraries Differ at the COOH Terminus from Those of Eluted Peptides. *J. Exp. Med.* **1997**, *185*, 367-371.
- [24] Ferrara, G. B.; Bacigalupo, A.; Lamparelli, T.; Lanino, E.; Delfino, L. *et al.* Bone marrow transplantation from unrelated donors: the impact of mismatches with substitutions at position 116 of the human leukocyte antigen class I heavy chain. *Blood* **2001**, *98*, 3150-3155.
- [25] Kawase, T.; Matsuo, K.; Kashiwase, K.; Inoko, H.; Saji, H.; *et al.* HLA mismatch combinations associated with decreased risk of relapse: implications for the molecular mechanism. *Blood* **2009**, *113*, 2851-2858.
- [26] Marino, S. R.; Lin, S.; Maiers, M.; Haagenson, M.; Spellman, S.; *et al.* Identification by random forest method of HLA class I amino acid substitutions associated with lower survival at day 100 in unrelated donor hematopoietic cell transplantation. *Bone Marrow Transplant.* **2012**, *47*, 217-226.
- [27] Bade-Döding, C.; Theodossis, A.; Gras, S.; Kjer-Nielsen, L.; Eiz-Vesper, B. *et al.* The impact of human leukocyte antigen (HLA) micropolymorphism on ligand specificity within the HLA-B*41 allotypic family. *Haematologica* **2011**, *96*, 110-118.
- [28] Qing, R. F.; Wileya, D. C. Structure of Human Histocompatibility Leukocyte Antigen (Hla)-Cw4, a Ligand for the Kir2d Natural Killer Cell Inhibitory Receptor. *J. Exp. Med.* **1999**, *190*, 113-123.
- [29] Rognan, D.; Zimmermann, N.; Jung, G.; Folkers, G. Molecular dynamics study of a complex between the human histocompatibility antigen HLA-A2 and the IMP58-66 nonapeptide from influenza virus matrix protein. *Eur. J. Biochem.* **1992**, *208*, 101-113.
- [30] Vita, R.; Overton, J. A.; Greenbaum, J. A.; Ponomarenko, J.; Clark, J. D.; *et al.* The immune epitope database (IEDB) 3.0. *Nucleic Acids Res.* **2015**, *43*, D405-D412.