JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

Subangstrom Accuracy in pHLA-I Modeling by Rosetta FlexPepDock Refinement Protocol

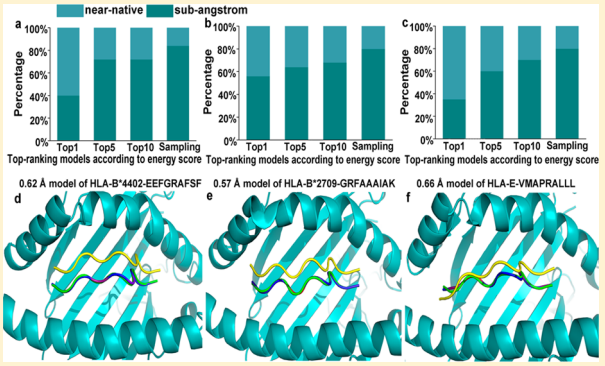
Tengfei Liu,[†] Xianchao Pan,[‡] Li Chao,[‡] Wen Tan,[‡] Sujun Qu,[‡] Li Yang,[†,‡] Bochu Wang,[†,‡] and Hu Mei*[,†,‡]

[†]Key Laboratory of Biorheological Science and Technology (Ministry of Education), Chongqing University, Chongqing 400044, China

[‡]College of Bioengineering, Chongqing University, Chongqing 400044, China

Supporting Information

ABSTRACT: Flexible peptides binding to human leukocyte antigen (HLA) play a key role in mediating human immune responses and are also involved in idiosyncratic adverse drug reactions according to recent research. However, the structural determinations of pHLA complexes remain challenging under the present conditions. In this paper, the performance of a new peptide docking method, namely FlexPepDock, was systematically investigated by a benchmark of 30 crystallized structures of peptide-HLA class I complexes. The docking results showed that the near-native pHLA-I models with peptide bb-RMSD less than 2 Å were ranked in the top 1 model for 100% (70/70) docking cases, and the subangstrom models with peptide bb-RMSD less than 1 Å were ranked in the top 5 lowest-energy models for 65.7% (46/70) docking cases. Furthermore, 10 out of 70 docking cases ranked the subangstrom all-atom models in the top 5 lowest-energy models. The results showed that the FlexPepDock can generate high-quality models of pHLA-I complexes and can be widely applied to pHLA-I modeling and mechanism research of peptide-mediated immune responses.



1. INTRODUCTION

Peptide-mediated interactions with proteins are vital for many cellular processes, such as signal transduction, regulatory networks, and immune recognition of living cells.[1,2] It has been estimated that nearly 40% of the protein−protein interactions are mediated by the flexible peptides that fold upon binding to protein receptors.[2] For the human immune system, peptide antigens bound to HLA molecules play a key role in T cell activation. In an antigen presenting cell (APC), the peptide-HLA (pHLA) complexes are presented and transported to the cell surface where T cell receptors (TCRs) recognize and examine peptide antigens resulting in specific T cell immune responses. Recent research indicates that drug hypersensitivity syndromes are also associated with pHLA molecules and T-cell receptors.[3,4]

Along with more and more structures of pHLA and pHLA-TCR complexes having been solved by X-ray and NMR experiments, great progress has been made in understanding the peptide-binding preferences and the mechanisms of pHLA-TCR interactions.[5−9] Nevertheless, experimental determinations of pHLA and pHLA-TCR structures remain challenging due to being expensive, time-consuming, and, in some cases, structurally unstable. Simultaneously, crystal structures are currently available for 32 HLA allotypes even though there are more than thousands of HLA allotypes.[10] Consequently, the development of pHLA modeling methods is essential for studying the T cell immune mechanism.

There are two key techniques in the modeling of peptide-HLA complexes:[11] (1) an effective conformational searching technique and (2) an accurate scoring function. The peptide-HLA interface contains huge amounts of rotatable bonds and large degrees of freedom that make it difficult to sample the conformers efficiently. On the other side, the available scoring functions often fail to rank native-like conformers or cannot distinguish them from decoys. In spite of these, several methods of peptide docking have been proposed during the past decade.

As a well-known peptide docking method, ZDOCK[12] performs a fast Fourier transform search of all possible binding modes of peptide/protein−protein complexes based on shape complementarity, desolvation energy, and electrostatics. It has been successfully applied to docking peptides to protein receptors, e.g., cysteine-rich intestinal protein 1 (CRIP1),[13] human epidermal growth factor 2 (HER2),[14] and human papilloma virus (HPV16) genetic variants.[15] Bordner et al.[16] docked peptides (8−10 mers) to HLA molecules by employing ICM (Internal Coordinate Mechanics) biased probability Monte Carlo search together with grid potentials. The peptide backbone-RMSD (bb-RMSD) of the docking poses obtained is

Received: April 28, 2014
Published: July 22, 2014

   dx.doi.org/10.1021/ci500393h | J. Chem. Inf. Model. 2014, 54, 2233−2242

Table 1. Results of Peptide Docking by Using a Peptide Template Bound to the Same HLA Serotype

	template		target		peptide RMSD			
serotype	pHLA	peptide sequence	pHLA	peptide sequence	start all-atom RMSD	start backbone-RMSD	top-5 backbone-RMSD[a]	top-5 all-atom RMSD[b]
HLA-A2	3MRB (A*0201)	NLVPMVHTV	1DUZ (A*0201)	LLFGYPVYN	3.72	3.44	0.78	2.00
			1TVB (A*0201)	ITDQVPFSV	3.64	3.22	1.08	2.41
			2V2W (A*0201)	SLYNTVATL	3.53	3.31	0.86	1.52
			3PWN (A*0201)	LLYGFVNYI	4.02	3.29	0.90	2.25
			4K7F (A*0201)	VCWGELMNL	3.72	3.43	0.62	1.11
HLA-B27	1KSN (B*2709)	GRFAAAIAK	1OGT (B*2705)	RRKWRRWHL	4.40	3.37	0.74	2.39
			1WOW (B*2709)	RRLPIFSRL	3.44	3.15	0.43	0.94
			2A83 (B*2705)	RRRWHRWRL	5.16	3.39	1.31	3.71
			2BST (B*2705)	SRYWAIRTR	4.13	3.23	0.77	2.98
			3BP4 (B*2705)	IRAAPPPLF	3.97	3.34	1.79	2.73
HLA-B35	2CIK (B*3501)	KPIVVLHGY	1A9E (B*3501)	LPPLDITPY	4.52	3.23	1.29	2.63
			3LKN (B*3501)	LPFERATIM	4.11	3.28	0.99	1.97
			3LKP (B*3501)	LPFDKSTIM	3.71	3.03	0.92	1.61
			3LKR (B*3501)	LPFERATVM	4.23	3.16	0.98	2.05
			3LKS (B*3501)	LPFEKSTVM	4.04	3.23	1.05	2.06
HLA-B44	3KPM (B*4402)	EEYLKAWTF	1M6O (B*4402)	EEFGRAFSF	3.89	3.41	0.62	0.92
			1SYS (B*4403)	EEPTVIKKY	4.45	3.30	0.86	2.30
			3KPP (B*4403)	EEYLQAFTY	4.52	3.43	1.46	2.80
			3L3J (B*4402)	EEAGAAFSF	3.88	3.33	1.20	1.58
			3L3K (B*4402)	EEFGAAASF	3.75	3.38	0.70	0.97
HLA-E	3BZF (E)	VMAPRALLL	1KPR (E)	VMAPRTVLL	3.46	3.23	0.49	0.98
			1KTL (E)	VTAPRTLLL	3.66	3.32	0.61	1.32
			2ESV (E)	VMAPRTLIL	4.06	3.51	0.73	1.17
			3AM8 (E)	VMGPRTLIL	3.57	3.14	0.63	1.17
			3BZE (E)	VMAPRTLFL	3.40	3.15	0.42	0.94

[a]The best peptide backbone RMSD within the top 5 models ranked according to the reweighted score. [b]The best peptide all-atom RMSD within the top 5 models ranked according to the reweighted score.

less than 2.5 Å. Based on annealing molecular dynamics, Niv et al.[17] proposed a PDZ-DocScheme protocol to dock peptides (4−8 mers) into PDZ domains. Good results were obtained with an average peptide bb-RMSD of 2.3 Å. Donsky et al.[18] developed a new tool, namely PepCrawler, allowing backbone flexibility of peptides and side-chain flexibility of both peptides and receptors. The modeling results showed that for the 88% of 25 complexes randomly selected from PeptiDB, one of the top three solutions obtained peptide interface bb-RMSD less than 1.6 Å. Antes et al.[19] applied DynaDock, a molecular dynamics-based method with a soft-core potential, to the docking studies over a small benchmark of 15 peptide−protein complexes, and obtained near-native and subangstrom models. Replica exchange all-atom discrete molecular dynamics (DMD) was also employed to simulate 10 well-characterized peptide−protein complexes (e.g., peptides binding to PDZ domain) and generated native-like (<2.5 Å) conformers.[20] Nevertheless, the universal applicability of this method requires further validation. In addition, 4 sequence-based methods,[21] namely DynaPred, NetMHC, SVMHC, and YKW, have been used to predict peptide binding affinities to MHC class I molecules. In most cases, sequence-based approaches are computationally more efficient than structure-based approaches. However, they are often limited by few experimental data available and the inability to provide structural information.

Rosetta is an excellent software suite that has the ability to model accurately macromolecular structures. Yanover et al.[22] applied Rosetta to construct position-specific binding energy matrices and used them to discriminate binding from nonbinding peptides for 11 HLA-A and 18 HLA-B subtypes. Good results were obtained with the median area under the receiver operating characteristic curve (auROC) of 0.756 and

dx.doi.org/10.1021/ci500393h | J. Chem. Inf. Model. 2014, 54, 2233−2242

Table 2. Results of Peptide Docking by Using Multiple Peptide Templates Bound to the Same HLA Serotype

	template		target		RMSD			
serotype	pHLA	peptide sequence	pHLA	peptide sequence	start all-atom RMSD	start backbone-RMSD	top-5 backbone-RMSD	top-5 all-atom RMSD
HLA-A2	1DUZ (A*0201)	LLFGYPVYN	3MRB (A*0201)	NLVPMVHTV	3.18	2.45	0.93	2.02
	1TVB (A*0201)	ITDQVPFSV			3.35	2.86	0.73	2.06
	2V2W (A*0201)	SLYNTVATL			3.75	3.66	0.73	1.79
	3PWN (A*0201)	LLYGFVNYI			3.20	2.68	0.76	1.97
	4K7F (A*0201)	VCWGELMNL			3.25	3.16	1.02	2.03
HLA-B27	1OGT (B*2705)	RRKWRRWHL	1K5N (B*2709)	GRFAAAIAK	3.01	2.11	0.57	0.69
	1WOW (B*2709)	RRLPIFSRL			3.69	3.38	0.62	0.91
	2A83 (B*2705)	RRRWHRWRL			3.77	2.81	1.29	1.61
	2BST (B*2705)	SRYWAIRTR			3.68	2.95	0.46	0.71
	3BP4 (B*2705)	IRAAPPPLF			3.83	2.80	1.55	1.88
HLA-B35	1A9E (B*3501)	LPPLDITPY	2CIK (B*3501)	KPIVVLHGY	3.63	2.25	1.14	2.03
	3LKN (B*3501)	LPFERATIM			3.13	2.12	1.39	2.60
	3LKP (B*3501)	LPFDKSTIM			3.10	2.10	1.29	2.53
	3LKR (B*3501)	LPFERATVM			3.47	2.54	1.37	2.71
	3LKS (B*3501)	LPFEKSTVM			3.60	2.76	1.23	2.50
HLA-B44	1M6O (B*4402)	EEFGRAFSF	3KPM (B*4402)	EEYLKAWTF	3.56	3.10	0.66	1.42
	1SYS (B*4403)	EEPTVIKKY			3.77	2.35	0.98	2.56
	3KPP (B*4403)	EEYLQAFTY			3.87	2.46	1.20	2.93
	3L3J (B*4402)	EEAGAAFSF			3.58	2.89	0.96	2.20
	3L3K (B*4402)	EEFGAAASF			3.30	2.74	0.75	1.49
HLA-E	1KPR (E)	VMAPRTVLL	3BZF (E)	VMAPRALLL	3.48	3.17	0.49	1.07
	1KTL (E)	VTAPRTLLL			3.34	3.10	0.54	1.14
	2ESV (E)	VMAPRTLIL			3.13	2.83	0.55	1.15
	3AM8 (E)	VMGPRTLIL			3.69	3.03	0.60	1.19
	3BZE (E)	VMAPRTLFL			3.39	2.53	0.61	1.02

0.786 for HLA-A and HLA-B, respectively. Patronov et al.[23] used AutoDock combined with RosettaDock to predict affinities of peptides binding to 4 HLA-DP subtypes. The Docking Score-based Quantitative Matrices (QS-QMs) derived from normalized free energy of binding (FEB) showed accurate prediction of peptide binding energies at an acidic environment. TCRFlexDock,[24] a flexible backbone docking protocol utilizing RosettaDock and ZRANK, was applied to model 20 TCR-pHLA complexes. The results showed that 80% of the docking cases generated near-native models among the top 10 lowest-energy models.

Recently, FlexPepDock, a peptide−protein docking method implemented in the Rosetta software, was proposed by Raveh et al.[25−27] There are two protocols provided by FlexPepDock: one is refinement and the other is *ab initio*. The refinement protocol of FlexPepDock[25] is intended for the cases where an approximate, coarse-grain model of the peptide−protein interaction is available. This protocol applies a Monte Carlo energy minimization approach to iteratively optimize peptide backbone and its rigid-body orientation and takes into comprehensive account side chain flexibility of both the peptide and protein receptor. FlexPepDock refinement protocol has been applied to an extensive and general benchmarking data set of 89 different peptide−protein complexes derived from the peptiDB data set[8] and obtained high-quality models that often are of subangstrom backbone level.

In order to investigate systematically the performance of FlexPepDock refinement protocol on pHLA modeling, a benchmark of 30 crystallized structures of pHLA-I was collected from the Protein Data Bank[28] (www.pdb.org). As we know, the starting conformation and orientation of a peptide is very important for a successful peptide-HLA docking. So, in this paper, two strategies for building peptide starting conformers were carefully investigated. The results showed that subangstrom models ranked in the top 5 lowest-energy models were generated for approximately 66% of our pHLA benchmark.

dx.doi.org/10.1021/ci500393h | J. Chem. Inf. Model. 2014, 54, 2233−2242

Table 3. Results of Peptide Docking by Using Multiple Peptide Templates Bound to Different HLA Serotypes

template		target		RMSD			
pHLA	peptide sequence	pHLA	peptide sequence	start all-atom RMSD	start backbone-RMSD	top-5 backbone-RMSD	top-5 all-atom RMSD
1K5N (B*2709)	GRFAAAIAK	3MRB (A*0201)	NLVPMVHTV	3.67	3.27	0.97	2.10
2CIK (B*3501)	KPIVVLHGY			3.52	3.01	0.93	2.34
3KPM (B*4402)	EEYLKAWTF			3.13	3.00	0.80	1.78
3BZF (E)	VMAPRALLL			3.35	3.23	0.45	1.87
3MRB (A*0201)	NLVPMVHTV	1K5N (B*2709)	GRFAAAIAK	3.88	3.08	0.84	0.97
2CIK (B*3501)	KPIVVLHGY			3.77	3.25	1.60	2.07
3KPM (B*4402)	EEYLKAWTF			3.86	3.35	1.20	1.51
3BZF (E)	VMAPRALLL			3.37	3.12	0.75	0.77
3MRB (A*0201)	NLVPMVHTV	2CIK (B*3501)	KPIVVLHGY	3.64	3.00	1.17	1.77
1K5N (B*2709)	GRFAAAIAK			4.04	3.18	1.42	2.52
3KPM (B*4402)	EEYLKAWTF			4.01	3.32	1.13	2.00
3BZF (E)	VMAPRALLL			3.71	3.05	1.36	2.25
3MRB (A*0201)	NLVPMVHTV	3KPM (B*4402)	EEYLKAWTF	4.08	3.25	0.86	1.75
1K5N (B*2709)	GRFAAAIAK			3.30	3.01	0.90	1.58
2CIK (B*3501)	KPIVVLHGY			4.32	3.39	0.94	2.40
3BZF (E)	VMAPRALLL			4.02	3.20	0.94	1.92
3MRB (A*0201)	NLVPMVHTV	3BZF (E)	VMAPRALLL	3.29	3.10	0.74	1.22
1K5N (B*2709)	GRFAAAIAK			3.84	3.38	1.53	1.81
2CIK (B*3501)	KPIVVLHGY			3.80	3.08	1.52	2.63
3KPM (B*4402)	EEYLKAWTF			3.68	3.24	0.66	1.09

2. METHODS

2.1. The Benchmark of the pHLA Complex. We searched pHLA-I complexes in the PDB database[28] in March 2014. To ensure the qualities of pHLA structures, those with resolution lower than 3.2 Å were under consideration. In most cases, the HLA-I molecules in the PDB database are cocrystallized with nonapeptides. So a total of 30 HLA-I molecules bound by nonapeptides were selected for the following docking studies. The selected HLAs belong to 5 serotypes, namely HLA-A2, HLA-B27, HLA-B35, HLA-B44, and HLA-E. Each serotype contains 6 HLA molecules cocrystallized with different nonapeptides (Table 1). More details of these 30 pHLA complexes can be found in the Supporting Information Table S1.

2.2. Strategies of Generating Peptide Starting Conformers. As we know, P2 and P9 anchor residues of the bound nonapeptides contribute, in a large degree, to sustain the overall shape of peptide backbones in the binding grooves of HLA-I molecules. Meanwhile, the distance between the alpha carbons of P2 and P9 anchors is relatively fixed for nonapeptides. For the 30 crystal structures of pHLA-I, the mean distance calculated is 18.63 Å with a standard deviation of 0.38 Å. Therefore, an approximate starting conformation of a target peptide can be templated by native peptides cocrystallized with HLAs.

In this paper, two strategies for generating peptide starting conformations were adopted: (1) The starting conformers of target peptides were derived from mutations of peptides bound to the same HLA serotypes; (2) The starting conformers were derived from mutations of peptides bound to different HLA serotypes.

In the first case, a template of pHLA complex for each HLA serotype was randomly selected, and the remaining 5 complexes in each serotype were used as target pHLA complexes (Table 1). Then, the starting conformers of the 5 target peptides were mutated from the peptide template selected. Afterward, the starting conformer obtained for each target peptide was orientated manually into the peptide binding groove of the corresponding HLA target. The resulting pHLA structures were then used as input files of FlexPepDock refinement, respectively.

Next, the template selected for each HLA serotype was conversely used as a target, of which the starting conformers were templated by the other 5 peptides, respectively (Table 2). For each HLA serotype, a total of 5 starting conformers for the target pHLA were then obtained and used as input files for FlexPepDock refinement, respectively.

In the second case, the 5 templates of pHLA complexes in Table 1 were employed as a target in turn (Table 3). Again, the starting conformers of target peptides were derived by mutations of the other 4 peptide templates. The resulting 4 starting conformers for each target were then used as input files for FlexPepDock refinement, respectively.

In sum, a total of 70 independent peptide dockings were performed by FlexPepDock refinement. The version of Rosetta FlexPepDock used in our study is 3.4. For the details of FlexPepDock protocol, please refer to the literature.[25−27]

2.3. The Flow of FlexPepDock Refinement. The outline of FlexPepDock refinement protocol is depicted in Figure 1.

2.3.1. Generating Input Files of pHLA Complexes. As we know, the P2 and P9 major anchor residues of nonapeptides get buried in pocket B and F of the HLA-I binding groove,

dx.doi.org/10.1021/ci500393h | J. Chem. Inf. Model. 2014, 54, 2233−2242

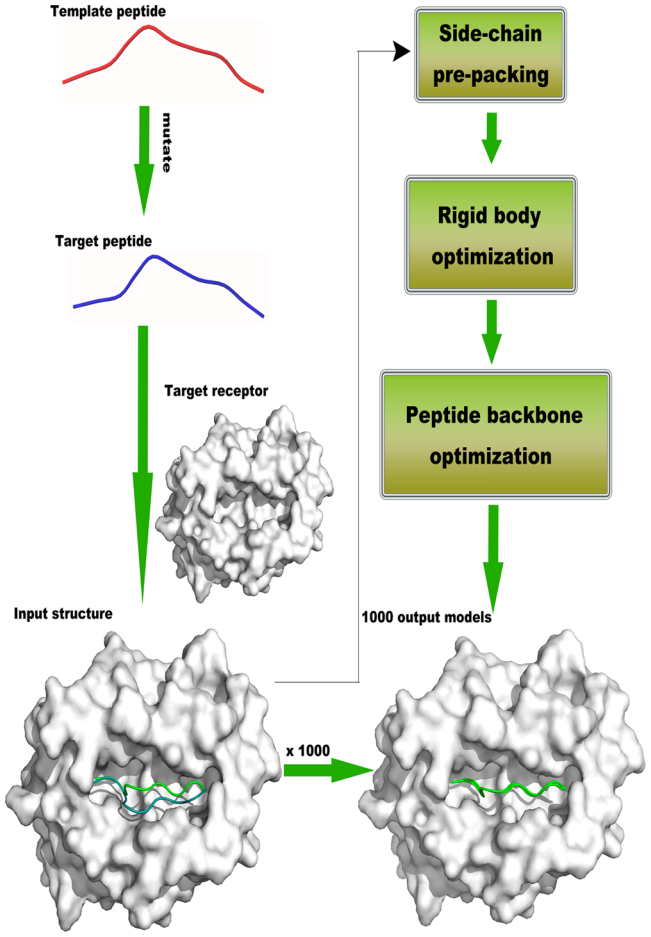


Figure 1. Outline of FlexPepDock refinement protocol.

respectively. Therefore, after starting conformations of target peptides were generated by mutating peptide templates in SYBYL 8.1,[29] the resulting peptide conformers were manually placed into the peptide binding groove of the target HLA molecule by a series of translation and rotation operations in rigid body,[30] while keeping the side chains of the P2 and P9 anchor residues pointing and close to B and F pockets, respectively. The merged peptide-HLA structures were then used as input files of FlexPepDock refinement protocol.

2.3.2. Prepacking the Side Chains of the Input Structures. It is necessary to prepack side-chains of each monomer before docking in order to remove internal clashes that are outside of the docking interface, as described in Raveh et al.[25] Herein, the number of output structures (nstruct) was set to 1. The prepack flag file is as follows:

$PATH_TO_EXE/FlexPepDocking.linuxgccrelease −database $PATH_TO_DB −s input.pdb −native native.pdb −ex1 −ex2aro −use_input_sc −flexpep_prepack −nstruct 1

2.3.3. Model Generation. The backbone of the prepacked peptide and its rigid-body orientation were first optimized relative to the receptor protein by employing the Monte Carlo Minimization approach, in addition to on-the-fly side-chain optimization.[25] Next, the structure obtained was refined in *n* independent FlexPepDock simulations, producing *n* candidate models. In order to effectively sample the conformational space, *n* was set to 1000, and the number of inner-cycles for both rigid-body and torsion-angle Monte Carlo Minimization was set to 10. Finally, the resulting models were ranked based on the Rosetta full-atom energy function. The following flag file was used:

$PATH_TO_EXE/FlexPepDocking.linuxgccrelease −database $PATH_TO_DB −s input.pdb **-mcm_cycles 10** −native native.pdb −pep_refine −ex1 −ex2aro −use_input_sc −nstruct 1000

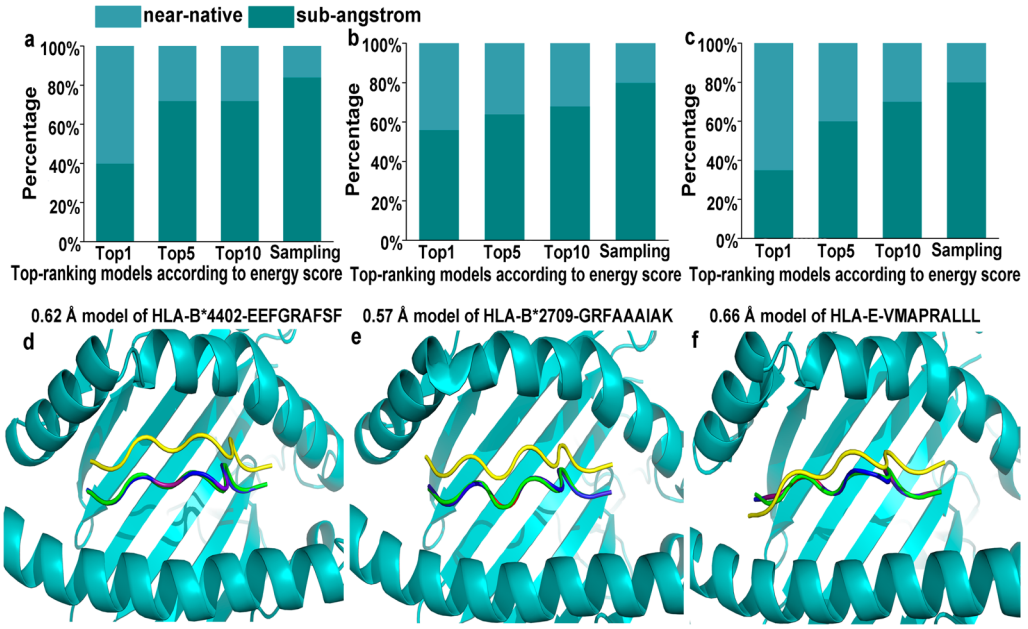


Figure 2. Sampling percentage of near-native and subangstrom models. Stacked bars represent the percent of subangstrom (dark cyan) and near-native models (palecyan). (a): docking peptides to HLA receptors by using one template bound to the same HLA serotype; (b): docking a peptide to a HLA receptor by using multiple templates bound to the same HLA serotype; (c): docking a peptide to a HLA receptor by using multiple templates bound to different HLA serotypes. The best model in the top 5 lowest-energy models: (d) 1M6O (template: 3KPM); (e) 1K5N (template: 1OGT); (f) 3BZF (template: 3KPM). Green: native conformer; Yellow: starting conformer. The peptides modeled are colored according to RMSD values: Blue: low RMSD, Red: high RMSD.

dx.doi.org/10.1021/ci500393h | J. Chem. Inf. Model. 2014, 54, 2233−2242

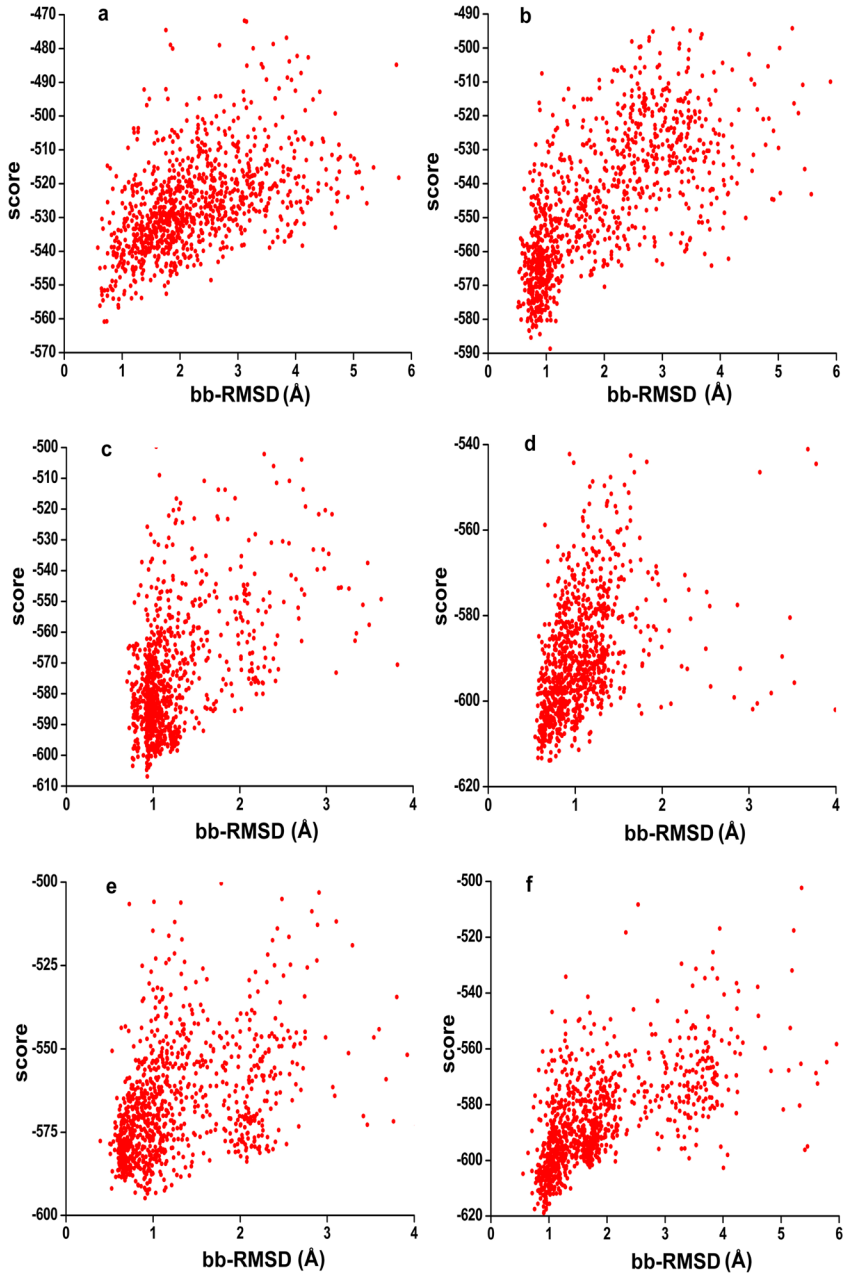


Figure 3. The plot of energy scores vs peptide bb-RMSD. The plots of energy landscape show the peptide bb-RMSDs (*x*-axis) of 1000 decoys versus reweighted energy scores (*y*-axis). (a) 4K7F (template: 3MRB); (b) 1OGT (template: 1K5N); (c) 3MRB (template: 3PWN) ; (d) 1K5N (template: 1OGT); (e) 3MRB (template: 3KPM); (f) 1K5N (template: 3BZF). (a), (b), (c), and (d) were templated by peptides bound to the same HLA serotype (Tables 1 and 2); (e) and (f) were templated by peptides bound to the different HLA serotypes (Table 3).

2.3.4. Assessment of Model Quality. According to a criteria defined by the CAPRI experiment (critical assessment of prediction of interactions),[31,32] the performance of peptide docking was evaluated by the backbone RMSD in the peptide−protein interface (bb-iRMSD) between a peptide model and the native peptide. A so-called near-native model is defined if the peptide bb-iRMSD is less than 2 Å, and a subangstrom model is defined if less than 1 Å. The interface residues of a peptide are defined as those whose C-Beta atoms (C-Alpha for Glycines) are within 8 Å away from any C-Beta atom of the partner protein. For pHLA-I complexes, all residues of peptides are located in the peptide-HLA interfaces. So, the RMSD and iRMSD have the same values in both backbone and all-atom levels. For clarity, bb-RMSD and all-atom RMSD were used in this paper.

2.3.5. The Energy Function Used for Model Evaluation. Score12 is the default full-atom energy function in Rosetta 3.4.[33] The total score of a model is computed as the weighted sum of different energy terms including van der Waals, solvation, hydrogen bonding, torsional, Coulombic, and harmonic restraints. However, the research of Raveh et al.[26] indicated that in most cases the reweighted-score (the weighted sum of score12, peptide score, and interface score, in which the interface score is given double weight, and the peptide score is given triple weight) performs better in predicting peptide binding affinities. The docking results of this paper also support this conclusion (Table S2, Supporting Information).

dx.doi.org/10.1021/ci500393h | J. Chem. Inf. Model. 2014, 54, 2233−2242

3. RESULTS AND DISCUSSION

3.1. High Quality pHLA Models, of Which Peptide Starting Conformers Were Templated by the Peptides Bound to the Same HLA Serotype. *3.1.1. Multitarget pHLA Modeling Templated by One Peptide Bound to the Same HLA Serotype.* The modeling results demonstrate the FlexPepDock refinement protocol is very promising with near-native models in the top 1 lowest-energy models for all 25 docking cases (Table 1). Subangstrom models are sampled in 84% cases (21/25). The refinement protocol reaches a 40% (10/25) success rate when ranking the subangstrom model as the top one lowest-energy model, and it reaches a 72% (18/25) success rate when ranking as one of the top 5 or top 10 models (Figure 2a). Among the docking cases shown in Table 1, the largest peptide bb-RMSD in the top 5 models is 1.79 Å. Further alignment studies prove that residues around the P5 position contribute heavily to this large deviation from the crystal structure.

For each HLA serotype, at least 3 docking cases reach the subangstrom level in the top 5 lowest-energy models (Table 1). For HLA-E serotype, all 5 cases yield subangstrom models in the top 5 models. Figure 2d shows the best subangstrom model of 1M6O in the top 5 models. It can be seen that the refinement protocol reproduces the crystal structure of the EEFGRAFSF bound to HLA-B*4402 quite well, of which the bb-RMSD and all-atom RMSD are 0.62 and 0.92 Å, respectively.

3.1.2. One-Target pHLA Modeling Templated by Multiple Peptides Bound to the Same HLA Serotype. Figure 2b shows that subangstrom models are sampled in 80% of all 25 cases (Table 2). 56% of the docking cases (14/25) achieve a subangstrom model in the top 1 lowest-energy model, and 64% (16/25) in the top 5 models, and 68% in the top 10 models (Figure 2b). Figure 2e shows the best subangstrom model of 1K5N (HLA-B*2709-GRFAAAIAK), of which the peptide bb-RMSD and all-atom RMSD are 0.57 and 0.69 Å, respectively.

In the modeling of the 2CIK complex templated by 5 different peptides (Table 2), the best peptide bb-RMSDs in the top 5 lowest-energy models are all larger than 1 Å. Alignments of the 5 models with the native pHLA structure show that the residue Val at position 5 has the largest RMSD.

In conclusion, the results above indicate that FlexPepDock refinement can, in most cases, achieve good performance on pHLA modeling when a peptide template bound to the same HLA serotype is available. In addition, RMSDs of residues around the P5 position are usually larger than that of other positions.

3.2. High Quality pHLA Models, of Which the Peptide Starting Conformers Were Templated by Peptides Bound to Different HLA Serotypes. As shown in Figure 2c, 35% docking cases (7/20) rank subangstrom models as the top 1 lowest-energy models, 60% cases (12/20) rank as the top 5, and 70% cases (14/20) as the top 10 models. Figure 2f shows the best subangstrom model of 3BZF (HLA-E-VMAPRALLL), of which the bb-RMSD and all-atom RMSD are 0.66 and 1.09 Å, respectively.

A slight drop in the success rate is observed in comparison with the docking cases templated by peptides bound to the same HLA serotype (Figure 2a, 2b, and 2c). Also, the best peptide bb-RMSDs in the top 5 lowest-energy models for 1K5N, 3KPM, and 3BZF (Table 3) are inferior to the corresponding results in Table 2. Again, no subangstrom model of the 2CIK complex is found in the top 5 lowest-energy models (Table 3). The reason remains a larger RMSD of Val at position 5.

In general, the FlexPepDock refinement is still an acceptable solution in the case where a peptide template bound to the same HLA serotype is not available.

3.3. Reweighted Scoring Function. As a rule of thumb, a successful scoring function combined with a successful search algorithm will generate score "funnels" where the lowest-energy models have low RMSDs.[34] In this paper, the reweighted score instead of the standard score12 was used for model evaluations. Figure 3 shows the representative plots of energy landscapes of the resulting pHLA models (see also Figure S1, Supporting Information). As expected, a trend of decreased energy scores is observed with the decreased RMSDs in most cases, which is characterized by the "funnel" shape.

Figure 4 shows the ability of the reweighted score to rank pHLA decoys. For the docking schemes in Tables 1, 2, and 3,

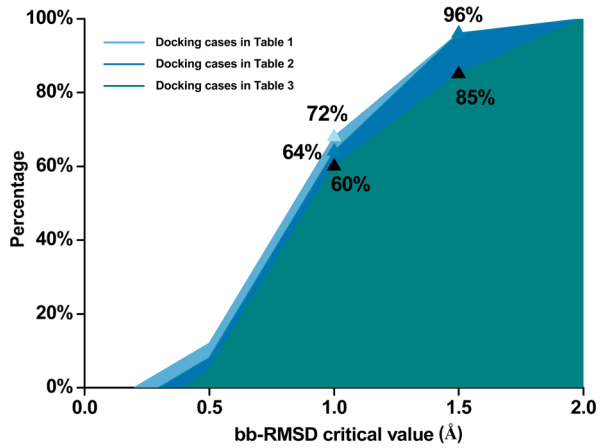


Figure 4. The ability of reweighted score to rank decoys. The plot shows the percentage of models (*y*-axis) with the bb-RMSDs smaller than critical values (*x*-axis) in the top 5 lowest-energy models.

the percentages of subangstrom models in the top 5 lowest-energy models reach 72%, 64%, and 60%, respectively, and the percentages of 1.5 Å models in the top 5 are 96%, 96%, and 85%, respectively. The results implicate that the reweighted score is qualified for peptide-HLA modeling.

3.4. RMSD of Major Anchor Residues of Peptide. For a nonapeptide bound to a HLA-I molecule, residues in positions 2 and 9 are major anchors, which play the most important roles in peptide-HLA interactions. In order to examine the ability of the FlexPepDock refinement protocol to model the major anchors, the average all-atom RMSD per residue of peptides were calculated for the top 5 lowest-energy models. As shown in Figure 5, the all-atom RMSDs of the 2 dominant anchors reach, in most cases, a subangstrom level (see also the Supporting Information Figure S2).

From another point of view, conserved interactions between anchor residues and HLA-I molecules make the conformational freedom of peptides be more focused on the center part of peptides, which commonly causes the central residues modeled with large deviations (Figure 5 and Figure S2, Supporting Information). To the best of our knowledge, residues in position 4−6 are generally extruded out of the peptide binding groove due to the steric hindrance and/or relatively weak

dx.doi.org/10.1021/ci500393h | J. Chem. Inf. Model. 2014, 54, 2233−2242

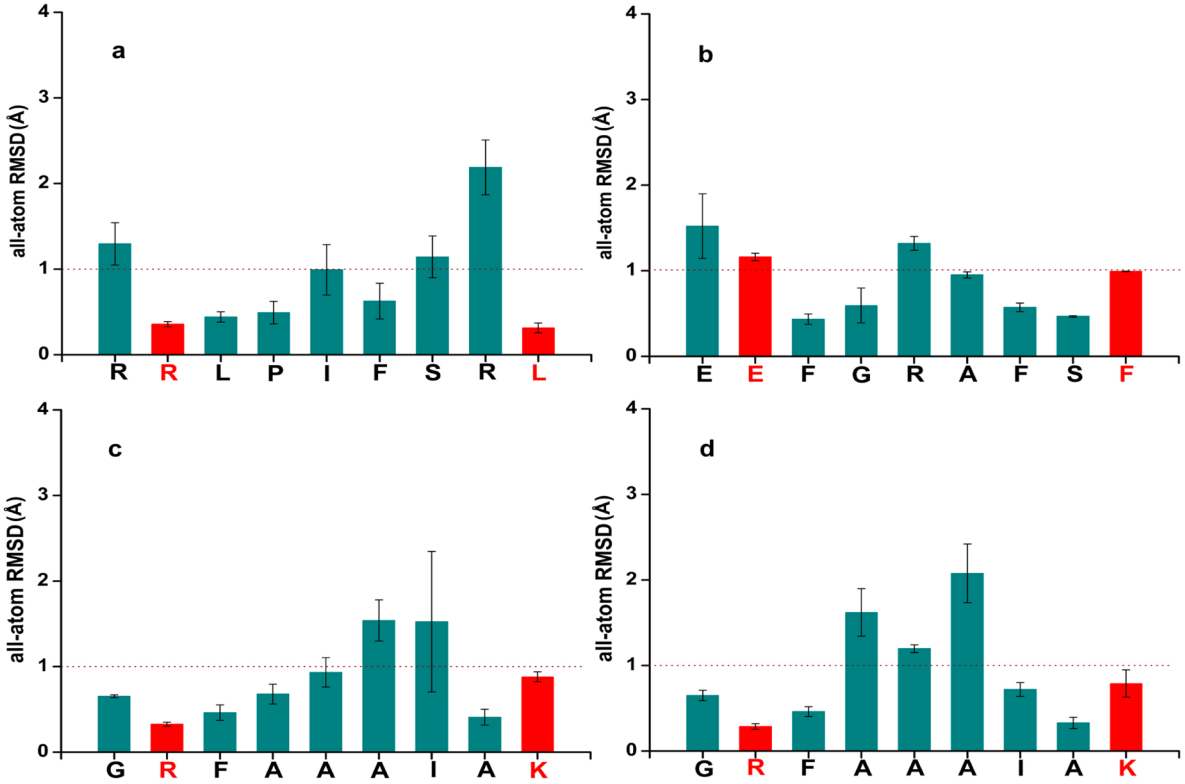


Figure 5. The all-atom RMSD per residue of peptide models. The stacked bars indicate the means of RMSDs in the top 5 lowest-energy models. The two major anchors are colored in red. (a) 1W0W (template: 1K5N). (b) 1M6O (template: 3KPM). (c) 1K5N (template: 1OGT). (d) 1K5N (template: 3BZF).

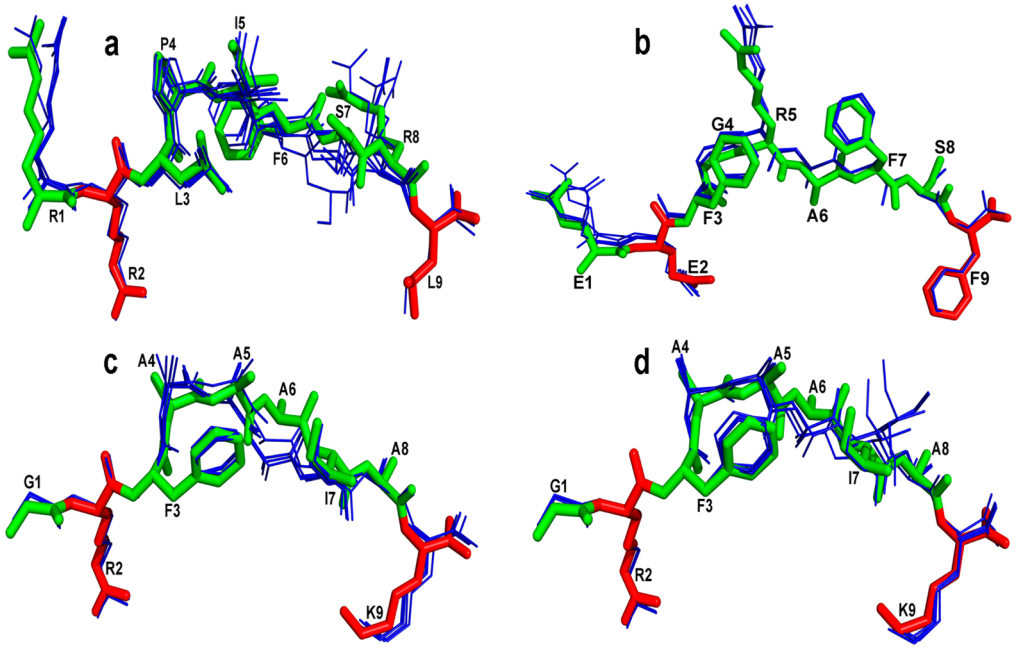


Figure 6. Alignment of the top 5 lowest-energy models with the native structures. The native peptide structures are shown in green sticks and the structures modeled in blue thin sticks. The two major anchors are highlighted in red. (a) 1W0W (template: 1K5N). (b) 1M6O (template: 3KPM). (c) 1K5N (template: 1OGT). (d) 1K5N (template: 3BZF). The structures of HLAs are not displayed for clear presentation.

interactions with HLAs. This also leads to large RMSDs in positions 4−6 of peptide in pHLA modeling.

3.5. The Achievements of FlexPepDock Refinement. In general, the FlexPepDock refinement protocol achieves good performance on pHLA modeling. The results show that a total of 46 docking cases (46/70, 65.7%) sample subangstrom bb-RMSD models in the top 5 lowest-energy models and that 10 docking cases (10/70, 14.3%) sample subangstrom all-atom models in the top 5 models. For the cases in Tables 1−3, the mean all-atom RMSDs for the decoys with subangstrom bb-

dx.doi.org/10.1021/ci500393h | J. Chem. Inf. Model. 2014, 54, 2233−2242

RMSD are 1.91 Å (±0.49), 1.68 Å (±0.41), and 1.71 Å (±0.32), respectively. Figure 6 shows the top 5 lowest-energy models of 1W0W, 1M6O, and 1K5N. It can be seen that the models generated are of high quality in an all-atom level.

3.6. pHLA Modeling in the Absence of Crystallized HLA Structures. It is often the case that no crystallized HLA structures are available for the modeling of pHLA complexes. On the consideration of the high structural conservation of HLA molecules, a starting conformation of a target pHLA complex can be prepared by direct mutating a pHLA template. Here, the pHLA-I complexes in Table 3 were employed to evaluate the docking performance of this strategy. The results showed that 70% (14/20) of the docking cases achieved subangstrom models in the top 5 lowest-energy models (Table 4). For 1K5N (template: 3KPM), subangstrom all-atom models were ranked among the top 5 models. Also, it can be seen that there is no significant difference between the docking results shown in Table 3 and Table 4.

Table 4. Results of Peptide Docking Derived from Direct Mutating pHLA Templates

template	target	RMSD			
pHLA	pHLA	start all-atom RMSD	start backbone-RMSD	top-5 backbone-RMSD	top-5 all-atom RMSD
1K5N (B*2709)	3MRB (A*0201)	2.03	1.01	0.88	1.99
2CIK (B*3501)		2.49	0.99	0.99	2.44
3KPM (B*4402)		1.36	0.67	0.67	1.26
3BZF (E)		1.91	1.14	0.81	1.28
3MRB (A*0201)	1K5N (B*2709)	2.55	1.01	0.82	1.83
2CIK (B*3501)		2.69	1.21	1.56	2.20
3KPM (B*4402)		2.47	0.86	0.49	0.64
3BZF (E)		2.68	1.62	0.99	1.38
3MRB (A*0201)	2CIK (B*3501)	3.07	0.99	0.84	1.82
1K5N (B*2709)		3.08	1.21	1.35	3.39
3KPM (B*4402)		3.26	1.03	1.55	2.46
3BZF (E)		3.22	1.59	1.04	1.98
3MRB (A*0201)	3KPM (B*4402)	2.13	0.67	0.70	1.59
1K5N (B*2709)		2.34	0.86	0.66	1.63
2CIK (B*3501)		3.2	1.03	0.99	2.28
3BZF (E)		3.11	1.41	1.05	2.22
3MRB (A*0201)	3BZF (E)	1.92	1.14	0.75	1.31
1K5N (B*2709)		2.41	1.62	0.85	1.42
2CIK (B*3501)		3.7	1.59	1.57	2.65
3KPM (B*4402)		2.55	1.41	0.91	1.66

4. CONCLUSIONS

There are a lot of computational methods which apply different strategies to predict peptide binding to HLA molecules. In general, sequence-based approaches are computationally more efficient than structure-based approaches and are often used to distinguish the binders from nonbinders with good predictive accuracies. However, sequence-based approaches cannot provide a straight structural interpretation of their results. The FlexPepDock, as one of structure-based methods, has the ability to generate a high-quality pHLA models and characterize the details in peptide-HLA interactions.

At present, however, energy scoring remains a key issue which needs to be better addressed in the structure-based methods. In this paper, although the reweighted score showed its ability to identify the subangstrom models, it failed to rank the native-like conformations in the top 5 lowest-energy models in some cases. It is technically feasible that for a particular modeling target, the weights of different energy terms can be manually adjusted to further improve ranking results.

Compared with docking of small molecules, peptide docking is more difficult and computationally complex due to more flexibility. In this paper, FlexPepDock refinement protocol is systematically investigated by a benchmark of 30 crystallized pHLA-I complexes. The results show that 100% docking cases rank near-native models in the top 1 lowest-energy models, and above 60% cases rank the subangstrom models in the top 5 models. The results indicate the FlexPepDock refinement protocol can address real world peptide-HLA docking problems and can be widely applied in pHLA modeling and research of peptide-mediated immune responses.

■ ASSOCIATED CONTENT

Supporting Information

Table S1: The details of the pHLA-I benchmark; Table S2: The comparison of scoring performance of Score12 vs Reweighted score; Figure S1: The plot of energy scores vs peptide bb-RMSD; Figure S2: The all-atom RMSD per residue of peptide models. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*Phone: +86-23-65102507. Fax: +86-23-65112677. E-mail: meihu@cqu.edu.cn.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This research was supported by the Natural Science Foundation of Chongqing (2013jcyjA10063), the Fundamental Research Funds for the Central Universities (CQDXWL-2012-121), the National Basic Research Program (973 Program) of China (2014CB541600), and the "111" project of "Introducing Talents of Discipline to Universities".

■ REFERENCES

(1) Pawson, T.; Nash, P. Assembly of cell regulatory systems through protein interaction domains. *Science* **2003**, *300*, 445−452.

(2) Petsalaki, E.; Russell, R. B. Peptide-mediated interactions in biological systems: new discoveries and applications. *Curr. Opin. Biotechnol.* **2008**, *19*, 344−350.

(3) Illing, P. T.; Vivian, J. P.; Dudek, N. L.; Kostenko, L.; Chen, Z.; Bharadwaj, M.; Miles, J. J.; Kjer-Nielsen, L.; Gras, S.; Williamson, N. A.; Burrows, S. R.; Purcell, A. W.; Rossjohn, J.; McCluskey, J. Immune self-reactivity triggered by drug-modified HLA-peptide repertoire. *Nature* **2012**, *486*, 554−558.

(4) Ostrov, D. A.; Grant, B. J.; Pompeu, Y. A.; Sidney, J.; Harndahl, M.; Southwood, S.; Oseroff, C.; Lu, S.; Jakoncic, J.; de Oliveira, C. A.; Yang, L.; Mei, H.; Shi, L.; Shabanowitz, J.; English, A. M.; Wriston, A.; Lucas, A.; Phillips, E.; Mallal, S.; Grey, H. M.; Sette, A.; Hunt, D. F.; Buus, S.; Peters, B. Drug hypersensitivity caused by alteration of the MHC-presented self-peptide repertoire. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 9959−9964.
(5) Hulsmeyer, M.; Hillig, R. C.; Volz, A.; Ruhl, M.; Schroder, W.; Saenger, W.; Ziegler, A.; Uchanska-Ziegler, B. HLA-B27 subtypes differentially associated with disease exhibit subtle structural alterations. *J. Biol. Chem.* **2002**, *277*, 47844−47853.
(6) Hoare, H. L.; Sullivan, L. C.; Clements, C. S.; Ely, L. K.; Beddoe, T.; Henderson, K. N.; Lin, J.; Reid, H. H.; Brooks, A. G.; Rossjohn, J. Subtle changes in peptide conformation profoundly affect recognition of the non-classical MHC class I molecule HLA-E by the CD94-NKG2 natural killer cell receptors. *J. Mol. Biol.* **2008**, *377*, 1297−1303.
(7) Macdonald, W. A.; Chen, Z. J.; Gras, S.; Archbold, J. K.; Tynan, F. E.; Clements, C. S.; Bharadwaj, M.; Kjer-Nielsen, L.; Saunders, P. M.; Wilce, M. C. J.; Crawford, F.; Stadinsky, B.; Jackson, D.; Brooks, A. G.; Purcell, A. W.; Kappler, J. W.; Burrows, S. R.; Rossjohn, J.; McCluskey, J. T Cell Allorecognition via Molecular Mimicry. *Immunity* **2009**, *31*, 897−908.
(8) London, N.; Movshovitz-Attias, D.; Schueler-Furman, O. The Structural Basis of Peptide-Protein Binding Strategies. *Structure* **2010**, *18*, 188−199.
(9) Vanhee, P.; Reumers, J.; Stricher, F.; Baeten, L.; Serrano, L.; Schymkowitz, J.; Rousseau, F. PepX: a structural database of non-redundant protein-peptide complexes. *Nucleic Acids. Res.* **2010**, *38*, D545−D551.
(10) Holdsworth, R.; Hurley, C. K.; Marsh, S. G. E.; Lau, M.; Noreen, H. J.; Kempenich, J. H.; Setterholm, M.; Maiers, M. The HLA dictionary 2008: a summary of HLA-A, -B, -C,-DRB1/3/4/5, and-DQB1 alleles and their association with serologically defined HLA-A, -B, -C, -DR, and -DQ antigens. *Tissue Antigens* **2009**, *73*, 95−170.
(11) Audie, J.; Swanson, J. Recent work in the development and application of protein-peptide docking. *Future Med. Chem.* **2012**, *4*, 1619−1644.
(12) Chen, R.; Li, L.; Weng, Z. ZDOCK: an initial-stage protein-docking algorithm. *Proteins* **2003**, *52*, 80−87.
(13) Hao, J.; Serohijos, A. W.; Newton, G.; Tassone, G.; Wang, Z.; Sgroi, D. C.; Dokholyan, N. V.; Basilion, J. P. Identification and rational redesign of peptide ligands to CRIP1, a novel biomarker for cancers. *PLoS Comput. Biol.* **2008**, *4*, e1000138.
(14) Banappagari, S.; Ronald, S.; Satyanarayanajois, S. D. A conformationally constrained peptidomimetic binds to the extracellular region of HER2 protein. *J. Biomol. Struct. Dyn.* **2010**, *28*, 289−308.
(15) Gokhale, P. S.; Sonawani, A.; Idicula-Thomas, S.; Kerkar, S.; Tongaonkar, H.; Mania-Pramanik, J. HPV16 E6 variants: Frequency, association with HPV types and in silico analysis of the identified novel variants. *J. Med. Virol.* **2014**, *86*, 968−974.
(16) Bordner, A. J.; Abagyan, R. Ab initio prediction of peptide-MHC binding geometry for diverse class I MHC allotypes. *Proteins* **2006**, *63*, 512−526.
(17) Niv, M. Y.; Weinstein, H. A flexible docking procedure for the exploration of peptide binding selectivity to known structures and homology models of PDZ domains. *J. Am. Chem. Soc.* **2005**, *127*, 14072−14079.
(18) Donsky, E.; Wolfson, H. J. PepCrawler: a fast RRT-based algorithm for high-resolution refinement and binding affinity estimation of peptide inhibitors. *Bioinformatics* **2011**, *27*, 2836−2842.
(19) Antes, I. DynaDock: A new molecular dynamics-based algorithm for protein-peptide docking including receptor flexibility. *Proteins* **2010**, *78*, 1084−1104.
(20) Dagliyan, O.; Proctor, E. A.; D'Auria, K. M.; Ding, F.; Dokholyan, N. V. Structural and Dynamic Determinants of Protein-Peptide Recognition. *Structure* **2011**, *19*, 1837−1845.
(21) Roomp, K.; Antes, I.; Lengauer, T. Predicting MHC class I epitopes in large datasets. *BMC Bioinf.* **2010**, *11*, 90.
(22) Yanover, C.; Bradley, P. Large-scale characterization of peptide-MHC binding landscapes with structural simulations. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 6981−6986.
(23) Patronov, A.; Dimitrov, I.; Flower, D. R.; Doytchinova, I. Peptide binding to HLA-DP proteins at pH 5.0 and pH 7.0: a quantitative molecular docking study. *BMC Struct. Biol.* **2012**, *12*, 20.
(24) Pierce, B. G.; Weng, Z. A flexible docking approach for prediction of T cell receptor-peptide-MHC complexes. *Protein Sci.* **2013**, *22*, 35−46.
(25) Raveh, B.; London, N.; Schueler-Furman, O. Sub-angstrom modeling of complexes between flexible peptides and globular proteins. *Proteins* **2010**, *78*, 2029−2040.
(26) Raveh, B.; London, N.; Zimmerman, L.; Schueler-Furman, O. Rosetta FlexPepDock *ab-initio*:Simultaneous Folding,Docking and Refinement of Peptides onto Their Recptors. *PLoS One* **2011**, *6*, e18934.
(27) London, N.; Raveh, B.; Cohen, E.; Fathi, G.; Schueler-Furman, O. Rosetta FlexPepDock web server-high resolution modeling of peptide-protein interactions. *Nucleic Acids Res.* **2011**, *39*, W249−W253.
(28) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235−242.
(29) Sybyl 8.1, Tripos Inc.: St. Louis, MO, USA, 2008. Available online: http://www.tripos.com (accessed January 26, 2011).
(30) Humphrey, W.; Dalke, A.; Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graphics* **1996**, *14*, 33−38.
(31) Janin, J.; Henrick, K.; Moult, J.; Eyck, L. T.; Sternberg, M. J.; Vajda, S.; Vakser, I.; Wodak, S. J. CAPRI: a Critical Assessment of PRedicted Interactions. *Proteins* **2003**, *52*, 2−9.
(32) Lensink, M. F.; Mendez, R.; Wodak, S. J. Docking and scoring protein complexes: CAPRI 3rd Edition. *Proteins* **2007**, *69*, 704−718.
(33) Rohl, C. A.; Strauss, C. E.; Misura, K. M.; Baker, D. Protein structure prediction using Rosetta. *Methods Enzymol.* **2004**, *383*, 66−93.
(34) Gray, J. J.; Moughon, S.; Wang, C.; Schueler-Furman, O.; Kuhlman, B.; Rohl, C. A.; Baker, D. Protein-protein docking with simultaneous optimization of rigid-body displacement and side-chain conformations. *J. Mol. Biol.* **2003**, *331*, 281−299.