Protein Science (1996), *5*:1272–1281. Cambridge University Press. Printed in the USA.

Free energy mapping of class I MHC molecules and structural determination of bound peptides

UGUR SEZERMAN, SANDOR VAJDA, AND CHARLES DELISI
Department of Biomedical Engineering, Boston University College of Engineering, Boston, Massachusetts 02218

(RECEIVED December 14, 1995; ACCEPTED April 17, 1996)

Abstract

Free energy maps of the binding site are constructed for class I major histocompatibility complex (MHC) proteins, by rotating and translating amino acid probes along the cleft, and performing a side-chain conformational search at each position. The free energy maps are used to determine favorable residue positions that are then combined to form docked peptide conformations. Because the generic backbone structural motif of peptides bound to class I MHC is known, the mapping is restricted to appropriate regions of the site, but allows for the sometimes substantial variations in backbone and side-chain conformations. In a test demonstrating the quality of predictions for a known MHC site using only a rotational and conformational search, we started from the crystal structure of the HIV-1 gp120/HLA-A2 complex, and predicted the HLA-A2 bound structures of peptides from the influenza matrix protein, the HIV-1 reverse transcriptase, and the human T cell leukemia virus. The calculated peptides are at 1.6, 1.3, and 1.4 Å all-atom RMSDs from their respective crystal structures (Madden DR, Garboczi DN, Wiley DC, 1993). A further test, which also included a local translational search, predicted structures across MHCs. In particular, we obtained the K^b/SEV-9 complex (Fremont DH et al., 1992, *Science 257*:919–927) starting with the complex between HLA-B27 and a generic peptide (Madden DR, Gorga JC, Strominger JL, Wiley DC, 1991, *Nature (Lond) 353*:321–325), with an all-atom RMSD of 1.2 Å, indicating that the docking procedure is essentially as effective for predictions across MHCs as it is for determinations within the same MHC, although at substantially greater computational cost. The requirements for further improvement in accuracy are identified and discussed briefly.

Keywords: free energy; grid search; homologous extension; major histocompatibility complex

The initial event in the cytotoxic T cell response to cellular infection is class I MHC recognition of proteolytically digested fragments of foreign proteins. The generic structure of the peptide–MHC complex is well understood, being based on the crystal structures of four different Class I MHC molecules, one of which, HLA-A2, has been solved with five different peptides (Madden et al., 1993). All structures show peptides bound in extended conformations with terminal residues closely aligned at both ends, but with considerable variability in backbone conformation and side-chain orientations in the central region of the peptide. Thus, for the four crystallized HLA-A2 nonamers, the C^α coordinates at positions 1, 2, 3, 8, and 9 are within 0.6 Å of one another, whereas the variation at position 6, where α-carbons reach their maximum variability, is 2.5 Å. Conformational variation between peptides that bind different MHCs is even greater; e.g., the C^α positions of residue 6 in peptides that bind K^b and HLA-A2 differ by 3.8 Å. A more extensive and more immediately important variation is in the orientation of the central side chains, which are highly sequence dependent, and show little consensus. Orientation is of particular interest because it determines which side chains are T cell accessible.

One of the major questions in computational structural biology is the extent to which a particular structure can be determined from a known generic structure. We are particularly interested in this for the class I peptide system, where the central problems are related to the docking and design of flexible peptides. The simplest structural questions that can be asked, and the ones addressed in this paper, are related to the accuracy with which the structure and binding affinity of nonhomologous peptides can be determined computationally. Because the general binding motif is known, the problem appears relatively uncomplicated. However, as we show below, the dispersion in side-chain orientations in the central region of the peptide, the mutual dependence of peptide and MHC side-chain conformations, and the required inclusion of solvation, all conspire against accurate all-atom predictions using traditional methods.

Our prime concern is the extent to which we can improve traditional approaches so that side-chain coordinates can be obtained reasonably accurately at relatively low cost. A traditional homologous extension approach, which will be used here for

Reprint requests to: Charles DeLisi, Department of Biomedical Engineering, Boston University College of Engineering, Boston, Massachusetts 02218; e-mail: delisi@buenga.bu.edu.

comparison, is mutation to produce the desired peptide, followed by a side-chain conformational search (SCCS). In the SCCS procedure, the backbone and the C^{α}-C^{β} bonds are fixed, and best side-chain conformations are searched about rotatable bonds (Bruccoleri & Novotny, 1992).

The search method we have developed falls within the general class of site mapping and fragment assembly procedures (Rosenfeld et al., 1993; Sezerman et al., 1993) and, in contrast to the SCCS method, does not assume that the backbone structure of the docked ligand is known. The first step of the search is the construction of free energy maps of the MHC binding site using separated amino acid residues as probes. Depending on the complexity of the problem, two different mapping procedures will be used. In the orientational search (OS), the C^{α} positions are set at the C^{α} position of the initial structure, the "probe" residue is rotated about an axis parallel to the long axis of the cleft, and, for each orientation, side-chain conformations are searched as in SCCS (see the Materials and methods). The second, more rigorous procedure also includes a translational search (TS), with translations of the C^{α} positions along a predefined grid, and an OS at every grid point. The free energy maps constructed by either the OS or TS methods provide favorable residue positions that are then concatenated to form the docked peptide conformations.

The site mapping and fragment assembly method was tested against three different Class I receptor systems, involving HLA-A2, H-2 K^b, and HLA-B27. The first of these has been crystallized with four different nonameric peptides: HTLV-1 Tax 11–19 (LLFGYPVYV); HIV-1 reverse transcriptase (RT) 309–317 (ILKGPVHGV); influenza virus matrix protein 58–66 (GILGF VFTL); and HIV-1 gp120 197–205 (TLTSCNTSV). We predict the structures of the first three peptides (referred to as HTLV-1, HIV-1 RT, and Inf) from the structure of HIV-1 gp120 using an OS in the mapping stage.

A more demanding application involves structures that have different bound peptide motifs (Falk et al., 1991; Rammensee et al., 1995). To explore a worst case scenario, we select HLA-B27, for which no single peptide structure is available, with a histone peptide constructed from the available consensus coordinates, and predict the structure of the (murine) K^b/SEV9 complex. The calculation is also performed in the opposite direction, i.e., predicting the HLA-B27/histone peptide complex from the K^b/SEV9 structure, although the results in this case are less revealing. The calculations are performed both without and with TS in the mapping. For the latter case, we find that the mapping/concatenation procedure, combined with a free energy target function that includes solvation, yields structures with all-atom RMSDs of 1.2 and 1.5 Å from those obtained crystallographically. We also show that, with one well-understood exception, the method effectively differentiates between T cell accessible and anchor side chains.

Results

HTLV-1 tax 11–19

The initial free energy evaluations for HTLV-1 amino acids eliminates position 6 as a possible anchor (Table 1). The combinations of favorable orientations for the seven remaining nonglycine side chains generate a total of $6 \times 4 \times 5 \times 1 \times 5 \times (-) \times 5 \times 5 \times 4 = 60{,}000$ structures, where (-) denotes a nonanchor residue, with no low-energy state identified by the mapping procedure. This number is reduced to 9,756 structures after all screening steps, and these are ranked according to their free energy. The lowest free energy structure had an all-atom RMSD from the crystal structure of 1.4 Å.

This result should be compared to that obtained by the traditional SCCS method. The best structure obtained by SCCS has an all-atom RMSD of 2.0 Å, with F3 and Y5 poorly predicted—albeit with somewhat better prediction of G4 (discussed below)—and a free energy 5 kcal/mol higher than that obtained when an OS is explicitly included (Table 2). Even with an explicit OS, however, the free energy of the crystal structure is more favorable by approximately 5 kcal/mol than that of our best predicted structure (−16 kcal/mol versus −11.1 kcal/mol).

HIV-1 reverse transcriptase 309–317

Local free energy mapping of the HIV-1 RT amino acids indicates that neither K3 nor P5 are likely to be a stabilizing peptide residue. The remaining side chains can be concatenated in $7 \times 4 \times (-) \times (-) \times (-) \times 2 \times 5 \times (-) \times 5 = 5{,}600$ ways. The imposition of bond angle constraints during concatenation reduced the number to 3,680 and, after minimization and elimination of structures with unacceptably high internal energy, the number remaining for free energy ranking is 1,206. The OS reduces the all-atom RMSD from 2.2 to 1.3 Å, markedly improving the determination of P5 and V6 (Table 3). The crystal free energy (−14.0 kcal/mol) is again much lower than that of the predicted structure (−8.0 kcal/mol).

Influenza virus matrix protein 58-66

The first screen eliminates F5 as a potential anchor and leaves $(-) \times 4 \times 7 \times (-) \times (-) \times 2 \times 4 \times 4 \times 4 = 3{,}584$ structures. After concatenation, the number of structures is reduced to 2,348, and these are further reduced to 989 by an internal energy screen. The OS does little to improve the accuracy (1.8 Å all-atom RMSD versus 1.6 Å all-atom RMSD in SCCS), largely due to not finding a good structure for F7, which deviates from the crystal structure by 2.7 Å (Table 4). We will return to this problem in the Discussion. Accuracy is again limited by the search procedure, the crystal structure having a free energy of −26.4 kcal/mol as opposed to −16.1 kcal/mol for the 1.4-Å predicted structure.

K^b/SEV 9

The initial free energy map of the K^b binding site eliminates N5, P7, and A8 as possible anchors. After free energy filtering, the number of possible concatenations was reduced to $6 \times 2 \times 1 \times (-) \times (-) \times 1 \times (-) \times (-) \times 4 = 48$. Eight of these violated bond angle constraints and were eliminated. Of the remaining 40, the best free energy structure (−14.7 kcal/mol), obtained without explicit allowance of central C^{α} translation, had an all-atom RMSD of 2.2 Å from the crystal structure (Table 5).

A major contributor to the relative inaccuracy of the prediction is Y6. The OS improves its RMSD from 6.3 Å (the result when only a conformational search is done) to 3.0 Å. Failure to improve further is a direct consequence of the initial 3.4-Å difference in the α-carbons at position 6. The result indicates that we have exceeded the limit of validity of the OS approxi-

Table 1. *Residue contribution to binding free energy in HLA-A2/HTLV-1 peptide complex for different orientations of the peptide residues*

Residue	Angle	E^{el}_{r-l} [a]	ΔG_h [b]	$T\Delta S_{sc}$ [c]	ΔG_{res} [d]	Residue	Angle	E^{el}_{r-l}	ΔG_h	$T\Delta S_{sc}$	ΔG_{res}
LEU1	0(47[e])	−4.0	−4.6	3.4	−5.2*	LEU2	0(5)	−4.6	−6.2	2.7	−8.1*
	30	−3.6	−3.5	3.3	−3.8		30	−4.5	−6.0	2.6	−7.9*
	60	−2.5	−3.7	3.1	−3.1		60	−0.6	−7.5	2.7	−5.4
	90	0.7	−4.9	2.4	−1.8		90	−2.8	−6.9	4.1	−5.6
	120	0.3	−5.1	2.4	−2.4		120	0.3	−6.4	4.2	−1.9
	150	−1.4	−5.9	2.8	−4.5		150	−0.3	−6.4	4.0	−2.7
	180	−4.4	−5.7	3.0	−7.1*		180	−0.9	−6.2	3.6	−3.5
	210	−4.8	−5.1	3.3	−6.6*		210	−2.4	−6.0	3.7	−4.7
	240	−4.0	−4.9	3.1	−5.8*		240	−3.0	−6.0	2.8	−6.2
	270	−2.0	−6.1	3.3	−4.8		270	−4.0	−6.3	2.7	−7.6*
	300	−3.4	−5.6	3.2	−5.8*		300	−3.9	−6.3	3.3	−6.9
	330	−4.0	−5.5	3.6	−5.9*		330	−6.0	−6.5	2.7	−9.8*
PHE3	0(202)	−3.6	−6.4	3.7	−6.3*	TYR5	0(230)	−4.1	−5.7	3.9	−5.9*
	30	−3.3	−7.7	3.4	−7.6*		30	−3.2	−6.7	4.4	−5.5*
	60	−3.8	−7.4	3.6	−7.6*		60	−3.0	−7.7	4.4	−6.3*
	90	−2.4	−7.8	3.5	−6.7*		90	−1.0	−6.5	4.4	−3.1
	120	−2.7	−8.5	3.4	−7.8*		120	−2.2	−5.6	3.6	−4.2*
	150	−2.0	−7.1	3.6	−5.5		150	−3.1	−3.9	4.0	−3.0
	180	−3.5	−5.9	3.9	−5.5		180	−2.0	−2.0	2.9	−1.1
	210	−3.6	−4.1	3.5	−4.2		210	−4.5	−4.0	3.3	−5.2*
	240	−3.4	−5.1	3.5	−5.0		240	−3.2	−2.5	2.9	−2.8
	270	−2.0	−4.0	3.7	−2.3		270	−3.0	−1.5	3.1	−1.4
	300	−2.1	−3.7	3.8	−2.0		300	−1.5	−1.2	2.9	0.2
	330	−3.3	−3.8	3.8	−3.3		330	−1.6	−3.0	3.1	−1.5
PRO6	0(56)	−1.9	−2.8	2.0	−2.7	VAL7	0(184)	−1.3	−1.9	2.5	−0.7
	30	−1.1	−2.6	2.3	−1.4		30	−3.2	−4.5	3.4	−4.3*
	60	−0.4	−1.8	2.1	−0.1		60	−3.4	−5.3	3.2	−5.5*
	90	−0.4	−1.8	2.2	0.0		90	−1.6	−6.2	3.5	−4.3*
	120	0.3	−2.8	2.3	−0.2		120	−2.2	−5.5	2.8	−4.9*
	150	0.1	−3.7	2.4	−1.2		150	−3.2	−5.4	2.9	−5.7*
	180	0.1	−4.1	2.2	−1.8		180	−3.0	−5.1	2.9	−5.2*
	210	−1.9	−4.7	2.0	−4.6		210	−2.7	−4.1	3.3	−3.5
	240	−1.8	−4.2	1.9	−4.1		240	–	–	–	–
	270	−1.9	−3.6	1.7	−3.8		270	−2.8	−3.4	3.2	−3.0
	300	−2.5	−3.2	1.7	−4.0		300	−2.0	−3.0	3.3	−1.7
	330	−2.6	−2.9	1.7	−3.8		330	−1.6	−3.2	3.3	−1.5
TYR8	0(22)	−4.6	−6.5	4.3	−6.8*	LEU9	0(290)	−4.4	−7.2	2.7	−8.9*
	30	−4.2	−5.5	4.3	−5.4*		30	−4.8	−7.0	2.7	−9.1*
	60	−2.4	−5.7	4.7	−3.4		60	−6.7	−7.0	2.5	−11.2*
	90	−4.0	−4.8	4.9	−3.9		90	−5.7	−4.2	3.6	−6.3
	120	−4.6	−5.3	4.9	−5.0*		120	−5.0	−4.6	3.2	−6.4
	150	−1.3	−5.7	4.7	−2.3		150	−5.1	−4.8	3.6	−6.3
	180	−1.4	−5.2	4.8	−1.8		180	−0.7	−5.0	3.7	−2.0
	210	−2.8	−6.7	4.9	−4.6		210	0.0	−5.5	3.7	−1.8
	240	−1.8	−6.3	4.1	−4.0		240	−2.3	−5.0	3.2	−4.1
	270	−3.2	−6.5	4.9	−4.8		270	−1.6	−4.9	3.1	−3.4
	300	−1.8	−9.8	4.5	−6.1*		300	−1.0	−6.0	2.8	−4.2
	330	−2.1	−8.2	4.6	−5.7*		330	−6.0	−6.9	3.1	−9.8*

[a] Electrostatic interaction energy.
[b] Hydrophobic conribution to the free energy.
[c] Free energy contribution due to sidechain entropy loss.
[d] Binding free energy contribution of the residue. Positions with * are retained for concatenation.
[e] 0 corresponds to 47° with the *x-z* plane.

Table 2. *All-atom and backbone RMSD and the solvent-accessible surface areas of predicted and observed htlv-1 peptide*

Residue	SCCS[a] aa_{RMSD}	OS[b] aa_{RMSD}	OS bb_{RMSD}	Observed area[c]	Pred. area
L1	1.0	0.7	0.7	28(S)	31(S)
L2	1.1	0.7	0.6	2(B)	2(B)
F3	2.2	1.5	0.8	2(B)	7(B)
G4	1.1	2.6	2.6	–	–
Y5	3.7	1.3	0.9	114(S), 38(OH)	127(S), 38(OH)
P6	1.4	1.5	1.8	45(S), 18(B)	46(S), 2(B)
V7	2.2	1.0	0.9	3(S)	12(S)
Y8	2.7	0.6	0.4	87(S), 28(OH), 6(B)	75(S), 22(OH), 6(B)
V9	1.0	1.1	0.3	2(B)	2(B)
Total	2.0	1.4			

[a] Side-chain conformations searched using the C^{α}-C^{β} orientation of the homologue.
[b] Orientations as well as conformations searched, as described in the Materials and methods.
[c] Solvent-accessible areas (Å^2) of side chain (S); backbone (B); and hydroxyl groups (OH).
aa_{RMSD}, all atom RMSD; bb_{RMSD}, backbone RMSD.

Table 3. *All-atom and backbone HIV-1 RT RMSD and the solvent-accessible surface areas of predicted and observed hivlrt peptide*[a]

Residue	SCCS aa_{RMSD}	OS aa_{RMSD}	OS bb_{RMSD}	Observed area	Pred. area
I1	1.4	1.2	0.8	36(S)	26(S)
L2	1.2	0.7	0.6	2(B)	4(B)
K3	1.2	1.3	0.8	5(S), 5(NZ)	27(S), 21(NZ)
E4	1.3	1.5	1.1	62(S), 27(B)	64(S), 21(B)
P5	3.4	1.9	1.7	76(S)	88(S)
V6	4.6	2.1	1.6	7(S), 28(B)	21(S), 35(B)
H7	1.5	0.8	0.8	41(S), 5(B)	45(S), 12(B)
G8	0.3	0.6	0.6	–	–
V9	1.1	1.0	0.3	2(B)	4(B)
Total	2.2	1.3	1.0	–	–

[a] See footnotes to Table 2.

Table 4. *All atom and backbone RMSD and the solvent-accessible surface areas of predicted and observed Inf peptide*[a]

Residue	SCCS aa_{RMSD}	OS aa_{RMSD}	OS bb_{RMSD}	Area (obs)	Area (pred)
G1	0.4	0.5	0.5	–	–
I2	1.3	1.2	0.7	4(B)	3(B)
L3	1.2	1.2	0.7	1(B)	4(B)
G4	1.4	0.8	0.9	–	–
F5	2.1	1.7	1.1	63(S)	42(S)
V6	2.6	1.9	1.2	48(S), 50(B)	85(S), 38(B)
F7	3.2	2.7	1.9	15(S)	72(S), 6(B)
T8	0.4	1.5	1.3	36(S), 6(B)	23(S), 13(B)
L9	0.9	2.1	1.0	0	2(B)
Total	1.8	1.6	1.0		

[a] See footnotes to Table 2.

Table 5. *All-atom and backbone RMSD and the solvent-accessible surface areas of predicted and observed of the SEV9 peptide*[a]

Residue	SCCS aa_{RMSD}	OS aa_{RMSD}	Full search[b] aa_{RMSD}	Full search bb_{RMSD}	Area (obs)	Area (pred)
F1	1.3	0.9	0.8	0.7	34(S)	29(S)
A2	1.0	0.7	0.6	0.6	–	–
P3	1.8	1.0	1.7	1.7	6(S), 1(B)	8(S)
G4	1.6	1.6	1.7	1.7	41(B)	53(B)
N5	3.7	3.7	1.7	1.7	88(S), 20(B)	64(S), 10(B)
Y6	6.3	3.0	0.7	1.0	7B	1(S), 6(B)
P7	2.0	3.6	1.0	0.9	7(S), 8(B)	1(B)
A8	0.8	1.0	1.0	0.8	46(S), 8(B)	61(S), 4(B)
L9	1.3	1.1	1.0	0.5	1(B)	7(B)
Total	2.7	2.2	1.2	1.2		

[a] See footnotes to Table 2.
[b] Includes explicit C^{α} translational search.

mation; viz, omission of explicit movement of the backbone. In this case, unlike homologous extensions between peptides bound to the same MHC, the prediction is from human to murine class I. The deviation in the central alpha carbons is too large to compensate by movement during concatenation (see the Materials and methods), and we must therefore treat it explicitly by a TS.

We thus extend the range of the calculation and, for the central residues (3–7), side-chain conformations and orientations are explored (i.e., an OS search is performed) at 125 grid points on a 5 × 5 × 5-Å cubic grid, centered at the B27 C^{α} positions. After orientational and free energy screening, the number of possible concatenations is 6 × 2 × 2 × (–) × 15 × 27 × (–) × (–) × 4 = 38,880. Bond angle constraints reduce the number to 18,036, each requiring a free energy evaluation. The computation time for minimizing and evaluating this final set of structures was approximately two months on an R8000 Silicon Graphics Challenger. This is probably more computation than required because all crystallized MHC peptides show similar terminal residue orientations. If residues 1 and 2 and 8 and 9 are fixed, the final number of concatenations is 801.

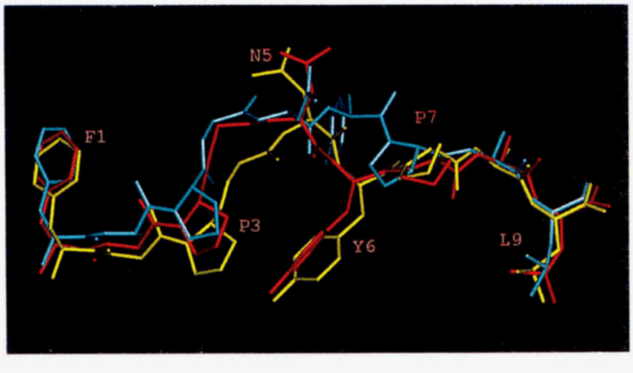

Fig. 1. Comparison of crystal, predicted, and mutated structures of SEV-9 peptide binding to K^b. Yellow is the crystal, red is the predicted, and blue is the mutated structure.

With a TS included, the lowest free energy structure found had Y6 and N5 at 0.7 and 1.7 Å from the crystal structure, and an all-atom RMSD of 1.2 Å (Fig. 1). A loss of accuracy at G4 occurs that reflects the limitations of grid size. This illustrates the trade-off between the cost of computation and the demand for accuracy, which we will discuss further. The all-atom RMSDs of the 15 best structures ranged from 1.2 to 1.8 Å (Table 6).

HLA-B27/histone peptide

Translational and orientational mapping found 4 positions for R2; 8 for I3; 16 for K4; 7 for T7, and 4 for L8. A5 and I6 were eliminated as potential anchors, and their possible structures

Table 6. *Binding free energies of the best 15 SEV9 complexes*[a]

	E^{el}_{r-l}	ΔG_h	$T\Delta S_{sc}$	Internal energy	ΔG_{pep}	aa_{RMSD}
1	−51.5	−27.7	17.4	41.9	−19.9	1.2
2	−51.2	−27.4	16.8	42.1	−19.7	1.5
3	−50.3	−28.0	16.6	42.1	−19.6	1.7
4	−52.4	−27.6	16.4	44.3	−19.3	1.5
5	−49.8	−28.4	16.8	42.2	−19.2	1.7
6	−50.5	−27.4	16.9	42.0	−19.0	1.7
7	−50.7	−27.4	16.8	42.7	−18.6	1.8
8	−51.2	−27.2	17.2	43.1	−18.1	1.6
9	−48.0	−27.3	16.7	40.6	−18.0	1.5
10	−49.0	−28.3	17.0	42.6	−17.7	1.4
11	−48.3	−28.8	16.7	42.9	−17.5	1.2
12	−47.8	−28.6	17.3	42.0	−17.1	1.6
13	−48.6	−28.0	17.6	42.8	−16.2	1.8
14	−47.2	−28.2	17.4	41.8	−16.2	1.7
15	−47.1	−28.2	16.8	42.4	−16.1	1.7
Obs[b]	−48.5	−28.5	16.2	38.8	−22.0	–

[a] See footnotes to Table 1.
[b] Crystal peptide with side-chain search.

were obtained by the IC BUILD routine in the CHARMm program (Brooks et al., 1983), as described in the Materials and methods.

Of the 14,336 concatenations, 10,121 were eliminated by bond angle filtering. The lowest free energy structure deviates from the crystal structure by 1.5 Å (Fig. 2; Table 7) and has a free energy of −62.6 kcal/mol; slightly lower than the free energy of the crystal structure (−60.2 kcal/mol). Of the 15 lowest free energy structures, 14 had RMSDs between 1.4 and 1.7 Å (Table 8). We recall that, for B27, the reference state is a structure determined with a mixture of endogenous peptides. Therefore, the comparison of predicted structures to the X-ray data is not as informative as for the other four peptides.

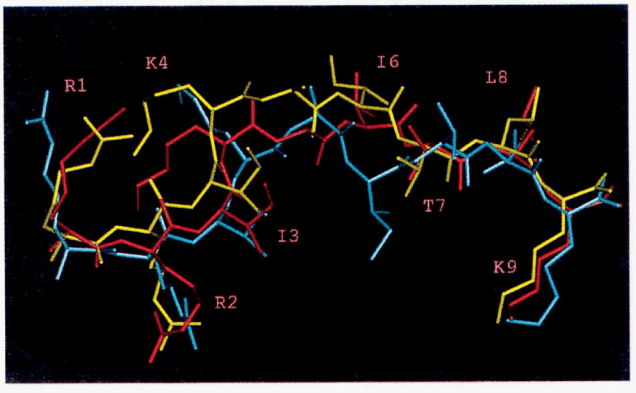

Fig. 2. Comparison of crystal, predicted, and mutated structure of histone peptide binding to HLA-B27. Yellow is the crystal, red is the predicted, and blue is the mutated structure.

Discussion

The questions of interest are related to (1) the consistency with which the best calculated free energy structure falls within a specified RMSD of the observed structure; (2) the extent to which anchor and T cell accessible side chains can be differentiated; (3) the extent to which the methods used here improve over standard methods; and (4) the main obstacles for further improvement and the likelihood that they can be overcome.

Predicting anchor and T cell recognition side chains

For the five peptides analyzed in this paper, the all-atom RMSDs between the observed structures and structures calculated using only SCCS are 2.0 (HTLV), 2.2 (HIV-1 RT), 1.8 (Inf), 2.6 (B27), and 2.7 Å (K^b). Adding a relatively coarse OS and an α-carbon TS for predictions across MHCs, with a primitive pruning procedure and a proper free energy function, yields all-atom RMSDs of 1.4, 1.3, 1.6, 1.5, and 1.2 Å, respectively. More importantly, all anchor and T cell accessible side chains are correctly predicted (Table 9).

If the goal is to use a known MHC–peptide complex to determine the structure of a peptide that binds the same MHC (such as HLA-A2/Inf from HLA-A2/HIV-1 gp120), the ability to correctly identify anchor residues is of no practical consequence, because they would be known. However, peptides that bind different MHCs (such as in HLA-B27 and K^b) generally have their anchor side chains at different positions. The ability to predict them opens the possibility of their determination for noncrystallized MHC molecules (the vast majority) and thus for finding binding segments, in any specified protein, for any specified MHC. This has clear implications for identifying candidates for peptide vaccines.

The results for all three HLA binding peptides indicate that if we did not know the key MHC binding residues, they could have been identified using an OS with a free energy target function. We also correctly distinguish all T cell accessible side chains from MHC buried side chains. The side chains of greatest interest are those toward the center, where most of the structural variability occurs. For HTLV-1, Y5 and Y8 are exposed in both the predicted and observed structures, and the OH group is available for hydrogen bonding with TcR side chains (Table 2). For HIV-1 RT, E4 and P5 are accessible; V6 and H7 are largely buried (Tables 3, 9). For K3, NH_3^+ is close to the solvent, thus minimizing the loss in desolvation energy, whereas the side chain itself, and hence its apolar groups, are largely buried in both

Table 7. *All-atom and backbone RMSD and the solvent-accessible surface areas of predicted and observed histone peptide*[a]

Residue	SCCS aa_{RMSD}	OS aa_{RMSD}	Full search[b] aa_{RMSD}	Full search bb_{RMSD}	Area (obs)	Area (pred)
R1	1.7	1.4	0.7	0.9	46(S)	31(S)
R2	1.9	1.7	1.5	0.7	–	–
I3	2.3	1.7	1.4	0.7	15(S), 1(B)	4(S)
K4	3.1	3.9	1.3	0.7	47(S), 37(B)	27(S), 30(B)
A5	2.4	2.8	1.8	0.8	11(S), 20(B)	20(S), 6(B)
I6	5.1	2.1	2.0	1.0	71(S), 30(B)	88(S), 19(B)
T7	2.1	1.3	1.9	0.8	9(S), 2(B)	46(S), 14(B)
L8	1.4	1.0	1.4	0.7	84(S), 4(B)	60(S), 7(B)
K9	1.0	0.9	0.7	0.6	2(B)	6(B)
Total	2.6	2.0	1.5	0.8		

[a] See footnotes to Table 2.
[b] Includes explicit C^α translational search.

Table 8. *Binding free energies of the best 15 histone peptide*[a]

Rank	E^{el}_{r-l}	ΔG_h	$T\Delta S_{sc}$	Internal energy	ΔG_{pep}	aa_{RMSD}
1	−90.4	−24.2	21.9	30.1	−62.6	1.5
2	−94.0	−24.0	20.9	36.1	−61.0	1.7
3	−93.8	−24.4	22.0	36.0	−60.2	1.7
4	−93.5	−23.8	21.8	35.9	−59.6	1.4
5	−90.1	−24.4	20.8	34.7	−59.0	1.7
6	−93.5	−22.9	21.9	36.1	−58.4	1.6
7	−94.1	−23.9	22.3	37.7	−58.0	1.7
8	−91.8	−23.4	22.1	36.6	−57.5	1.6
9	−91.1	−23.7	22.2	35.2	−57.4	1.5
10	−91.2	−24.7	24.9	33.8	−57.2	1.5
11	−89.3	−21.4	21.0	32.9	−56.8	1.7
12	−88.8	−21.6	21.4	33.9	−55.1	1.6
13	−87.6	−24.1	26.6	30.3	−54.8	2.6
14	−87.2	−23.6	22.5	33.7	−54.6	1.6
15	−83.8	−22.0	21.1	30.2	−54.5	1.7
SM[b]	−78.9	−24.5	27.6	25.4	−50.4	2.6
Obs[c]	−81.1	−24.2	23.9	21.2	−60.2	–
Obs[d]	−86.8	−24.1	21.5	24.5	−64.9	–

[a] See footnotes to Table 1.
[b] Histone peptide obtained by side-chain mutation.
[c] Crystal peptide with side-chain search.
[d] Crystal peptide with predicted R1 and K4 configurations.

Table 9. *Percent solvent-accessible surface area*

INF	G	I	L	G	F	V	F	T	L
Obs[a]	–	2	0	–	48	48	10	26	0
Pred[b]	–	2	1	–	36	67	0	28	0
HIV 1-RT	I	L	K	E	P	V	H	G	V
Obs	15	0	3	55	78	18	22	–	0
Pred	11	0	15	53	91	37	27	–	0
HTLV	L	L	F	G	Y	P	V	Y	V
Obs	14	0	1	–	75	61	3	43	0
Pred	14	0	2	–	86	47	12	40	0
SEV	F	A	P	G	N	Y	P	A	L
Obs	13	0	7	–	68	4	12	53	0
Pred	11	0	9	–	53	6	1	64	1
Histone	R	R	I	K	A8	I	T	L	K
Obs	22	0	12	59	52	50	10	64	1
Pred	16	0	3	40	47	53	38	49	2

[a] Surface area of crystal structure.
[b] Predicted surface area.

predicted and observed structures, with 15% and 3% solvent-accessible surface areas, respectively. The potentially protonated nitrogen in H7 has 14% of its area exposed in the crystal structure and 15% in the predicted. For His in its free form, 21% of the N surface area is exposed. Consequently, both the crystal and calculated structures predict that this group should be available for T cell receptor (TcR) interaction in its free form.

For the Inf peptide, the agreement between predicted and observed groups is similarly good. There is some inaccuracy in V6, which is predicted to be 67% solvent accessible, compared to the observed value of 48%. In either case, it is possible that V6 interacts favorably with a TcR hydrophobic region. For comparison, the orientation of Tyr 5 in the HTLV-1 peptide is completely mispredicted using only SCCS (4.4 Å RMSD), whereas the additional OS obtains it at an RMSD of 1.2 Å. Similarly, for the HIV-1 RT peptide, the orientations predicted for Pro 5 and Val 6 using only SCCS differ by 3.4 and 3.9 Å from the observed, whereas OS finds them to within 2.0 and 2.1 Å, respectively.

A more demanding test of the method is the ability to predict across MHC molecules, where the peptide anchors are not at corresponding positions. The conserved side chains of peptides bound to the B27 class I molecule are at positions 2 and 9, whereas the conserved side chains of peptides bound to the murine K^b molecule are at positions 6 and 9. Position 6 in the B-27 peptide therefore provides little information about position 6 in the K^b peptide (SEV9). Moreover, the current X-ray structure of B-27 is based on a mixture of endogenous peptides, and some of the peptide side chains are not determined by direct observation. This makes the predicted 1.2-Å all-atom RMSD K^b structure all the more encouraging. The lack of a highly resolved B-27 peptide might also explain some of the apparent anomaly, at least with respect to the other systems, of finding a predicted peptide with a free energy as low as that of the crystal peptide.

Near-term prospects for increased accuracy

The significance of these results lies not simply in their degree of accuracy and range of improvement over a standard side-chain mutation and energy-minimization calculation, but in what they tell us about current obstacles for further near-term increases in accuracy. We will briefly review how various features of the algorithm and those of the target function can be improved to provide more accurate predictions.

Translation

The most direct way to improve the search is to include translation routinely. The philosophy in this paper was, in part, to explore the trade-off between speed and accuracy. Because α-carbons translate by as much as 1 Å during concatenation, omission of explicit translation for homologous extensions between HLA binding peptides seemed reasonable, and the limited search yields good results in most cases. That the restriction nevertheless does impair accuracy is suggested by our ability to obtain more accurate predictions for K^b starting from B-27 when explicit translations are allowed. Indeed, for the SEV-9 peptide, the RMSD was reduced from 2.2 to 1.2 Å, albeit at a two order of magnitude increase in computer time.

The expected gain for extensions within HLA-A2 will not be this large because the accuracy is already high. Whether the gain is worth the additional cost depends on objectives. Nevertheless, the limitation due to fixed C^α positions is seen by the inaccuracy of G4 (2.6 Å) for the HTLV-1 peptide. An equally important indicator of this limitation is the large discrepancy between the predicted and calculated free energies of the flu peptide. The C^α for V6 differs from the C^α at position 6 in HIV-1 gp by 2.0 Å, causing a 1.3-Å RMSD at this position in the final structure. This increases the energy of the predicted structure by 3.1 kcal/mol. In addition, the error propagates to F7 and T8,

Table 10. *Binding free energies of the best 10 Inf phe7 side-chain configurations*

	Congen energy	Free energy	aa_{RMSD}
1	20.9	−6.0	2.1
2	25.1	−5.8	2.4
3	27.2	−5.8	2.5
4	29.3	−4.8	2.2
5	32.7	−6.2	1.1
6	33.6	−4.8	2.3
7	35.2	−7.1	0.3
8	39.9	−5.1	2.5
9	43.6	−6.6	0.4
10	48.4	−4.6	2.7

increasing the energy relative to the crystal by an additional 2.6 kcal/mol.

Energy selection

At present, energy, rather than free energy is used to select the best structure at each orientation. The 12 structures thus selected for each α carbon are then ranked according to their free energies. The result for Inf F7 illustrates that using energy can lead to inaccurate predictions. Although the C^{α}-C^{β} bond orientation is determined correctly, the ring is flipped up, away from the floor of the cleft, rather than toward it as in the crystal. The source of the difficulty has been identified by looking at several conformations with the same C^{α}-C^{β} bond orientation, ranked according to their CHARMm energy (Table 10). Results clearly show that the error is in using the CHARMm energy to select the best structure at a given orientation. The conformation with the lowest free energy (0.3 Å RMSD) is some 15 kcal/mol above the lowest CHARMm energy structure. The failure thus arises from introducing free energy too late in the process. In order to correct it in a general way, we would need to retain, temporarily, several conformations at each orientation, and then rank them according to their free energy. We could then continue to use a winner-take-all strategy for retaining a single conformation at each orientation. There is no substantial difficulty in doing this in general. When, in fact, we use the best free energy structure in the buildup, the final conformation of Phe 7 has an RMSD of 1.9 Å, and the overall RMSD of the flu peptide is 1.4 Å (Table 11) rather than 1.6 Å.

Energy is used in the first instance for several reasons, the most important being the assumed cancellation of solute-solute and solute-solvent van der Waals effects. The assumption obviates the need for including van der Waals evaluation in the free energy function, because it is the free energy difference that matters. At the same time, however, the lack of explicit van der Waals interactions precludes the possibility of direct minimization. A second, more practical reason for using energy as a first pass selection criterion is that the calculation of the solvation term, and particularly of its derivative, would be computationally very demanding. These difficulties can, in fact, be circumvented by new approaches to solvation that are under development in our laboratory (Zhang, in prep.).

We have performed additional calculations to study the effect of retaining only a number of low-energy residue orientations (see Table 1), and then screening on the basis of energy, rather than free energy, in the concatenation stage. To this end, starting from the best structure found for the Inf peptide, we generated, retained, and combined all of the 12 rotational positions for the side chains of residues 3, 5, 6, and 7 (recall that the Inf peptide has a Gly at position 4, the terminal positions are almost invariant, and the backbone at positions 2 and 8 also shows very limited variability). Exploiting these constraints results in $12^4 = 20{,}736$ conformations that were generated and subjected to local minimization as described in the Materials and methods. We have evaluated the free energies and the RMSD values without any pruning. The extended search slightly improved the results, decreasing the free energy to −17.2 kcal/mol (from −16.1 kcal/mol), and the RMSD to 1.3 Å (from 1.4 Å). Notice that the extended search is computationally feasible only with five residues fixed, and the resulting small improvement suggests that energy-based pruning is an acceptable way of reducing the computational burden without deleterious effect on the prediction.

MHC side chains

The procedure used in this paper locally minimizes all MHC side chains in the cleft that interact with peptide. For extensions

Table 11. *All-atom RMSD and the solvent-accessible surface areas of predicted and observed Inf peptide*[a]

Residue	SCCS aa_{RMSD}	OS aa_{RMSD}	Area (obs)	Area (pred)
G1	0.4	0.7	–	–
I2	1.3	1.2	4(B)	3(B)
L3	1.2	1.2	1(B)	3(B)
G4	1.4	0.7	–	–
F5	2.1	1.6	63(S)	46(S)
V6	2.6	1.8	48(S), 25(B)(O)	82(S), 16(B)(O)
F7	3.2	1.9	15(S)	0
T8	0.4	0.9	6(B), 36(S)	13(B), 36(S)
L9	0.9	1.6	0	3(B)
Total	1.8	1.4		

[a] Using best free energy Phe 7 conformation.

within the same MHC, there are up to seven side chains (positions 73, 95, 97, 116, 156, 163, and 167) whose conformations are dependent on the bound peptide. Some or all of these must be searched extensively, either by a Monte Carlo method, or an exhaustive grid search, as was used for finding peptide conformations. Extensions across MHC molecules may require searching as many as 19 MHC side chains, and may not be possible on a routine basis. For example, 152E and 156L in K^b are in error by 2.3 Å and 1.8 Å, repectively, affecting the accuracy of prediction for N5 in the SEV9 peptide. Conformational search of residues 152E and 156L along with N5 would increase the accuracy of the prediction.

The target function

The calculated free energy of the observed structure is lower than that of the calculated structure for four of the five peptides (Table 12) and, in three cases, the difference is substantial, suggesting that the search procedure is currently the primary limitation on accuracy. The exception is the histone peptide, which has a predicted structure free energy about 4% lower than that of the crystal structure. However, the crystal structure in this case is not well defined.

Materials and methods

Free energy target function

The free energy function and its validation have been described in detail previously (Vajda et al., 1994). Briefly, the free energy target function consists of three terms: the electrostatic interaction energy between the ligand and the receptor; the solvation free energy, and the side-chain conformational entropy. The electrostatic energy (v. 2.0 of the CONGEN program with CHARMm 19 parameters) is calculated using a distance-dependent dielectric constant of $4r$. Solvation free energy and side-chain entropy loss are calculated using an extended version of the program by Eisenberg and McLachlan (Vajda et al., 1994).

The binding free energy is the difference between the average free energy of the free peptide (which is conformationally flexible), and the free energy of the bound peptide. The development assumes that stress-free equilibrated structures have comparable van der Waals interactions (Novotny et al., 1989; Krystek et al., 1993); i.e., any protein-protein van der Waals interaction lost (gained) by structural modification are compensated for by van der Waals interaction gained (lost) by the accompanying changes in solvation (Lee & Richards, 1991).

Because we are interested in the bound structure associated with the minimum binding free energy, we need only rank the free energy changes, without great concern for their absolute values. This ranking need involve only the differences in free energies between different conformations of the same peptide MHC complex; i.e., the free energy of the free peptide can be ignored. We thus take the free energy of the complex as the target function in our calculations.

Table 12. *Free energies of observed and predicted peptide-MHC complexes*

	HTLV-1	HIV1-RT	Inf	Histone	SEV9
Observed	−16.0	−14.0	−26.4	−60.2	−22.0
Predicted	−11.1	−8.0	−16.1	−62.6	−19.9
RMSD	1.4	1.3	1.4	1.5	1.2

Mapping

The target function for each amino acid, considered independently, is calculated for a range of translational, orientational, and conformational states compatible with the presence of the MHC and with observed ranges in the locations of amino acids in corresponding positions in other MHC-bound peptides.

For peptides bound to the same MHC, the observed variation in alpha carbon location is comparable to the change that occurs during the peptide concatenation procedure, as explained below. Therefore, the C^α position for exploring side-chain orientation is set at the C^α position of the initial structure. For peptides bound to different MHCs, corresponding α-carbons can differ by up to 4.8 Å, and we therefore include translational motion.

The orientational space of each peptide residue is searched by 30° rotations of the C^α–C^β bond about an axis parallel to the long axis of the cleft (Cornette et al., 1993). The orientational angle is zero along the floor of the cleft, and it increases as the vector rotates from the $\alpha 1$ toward the $\alpha 2$ helix. For K^b and B-27, the orientational/conformational searches were conducted at each C^α position in a 125-$Å^3$ volume, using a 1-Å grid spacing for the central residues (4–7).

For each orientation, side-chain conformations are searched exhaustively in 15° increments about rotatable bonds, using the CONGEN program (Bruccoleri & Novotny, 1992) with the CHARMm 19 energy function. All variable MHC side chains having any atom within 4 Å were also searched exhaustively (SCCS). Only the lowest-energy conformations are retained at each orientation.

This winner-takes-all strategy is the first of the screening steps, and leaves us with 12 conformational states for each side chain, one for each orientation. The states are ranked according to free energy. The 12 states are then further pruned by retaining only those within 30% of the lowest free energy. At this stage, we generally have fewer than six states per side chain, as shown in Table 1 for the HTLV peptide. Angles are incremented by 30°, starting with the orientation in the crystal. Thus, LEU1 is at 47° with respect to the Z axis defined in the text, and subsequent entries are at 77°, 107°, etc.

Prior to free energy evaluation, all structures are refined by performing local minimization using the CHARMm energy function until the change in energy settles to less than 0.1 kcal/step, averaged over the preceding 10 steps. This generally requires about 80 steps of minimization and provides structures that are free of steric clashes (Brooks et al., 1983).

This procedure is applied to every side chain in the sequence, but not all of them are found to have strong binding potential for the MHC. We define these weakly interacting side chains as those having free energies higher than the least favorable anchors. Known anchors range in free energy from a low of −13.7 kcal/mol (R2 of the B27 bound histone peptide) to a high of −5.7 kcal/mol (Y6 of SEV9). For side chains with free energies above −5 kcal/mol, we adopt the physically realistic viewpoint

that their conformation and orientation is determined by bond and angle constraints imposed by the states of the adjoining peptide side chains, and by van der Waals constraints imposed by the MHC environment. When a segment consists of a consecutive sequence of three or more side chains that fall into the weak binding category, its backbone conformation is not uniquely determined by standard polypeptide geometry, and has to be calculated by loop closure methods (Bruccoleri & Novotny, 1992; Zheng et al., 1993; Vasmatzis et al., 1994). For HLA-A2 binding peptides, the number of consecutive side chains that fall into the weak binding category never exceeds two. Although determining the missing backbone does not involve a conformational search, introducing some variation in bond lengths and valence angles may be necessary. The allowance of such variation is, in fact, how the standard routine in CHARMm (IC BUILD) builds missing coordinates. The result is generally a highly stressed segment, which is subsequently relaxed by energy minimization.

The state of each amino acid is then selected from this free energy map in a way that generates a set of low free energy structures for the peptide as a whole. There are a number of possible ways in which low free energy structures can be composed from mapping data (Rotstein & Murcko, 1993), including Monte Carlo methods (Goodsell & Olson, 1990; Yue, 1990; Caflisch et al., 1992; Hart & Read, 1992), dynamic programming algorithms (Gulukota et al., 1996), and multiple copy algorithms (Miranker & Karplus, 1991; Caflisch et al., 1993; Rosenfeld et al., 1993). In this paper, we pursue an approximate approach aimed at simplicity and physically motivated pruning strategies.

Concatenation

All combinations of orientations of the stabilizing residues are concatenated, including those determined by loop closure, eliminating only structures whose bond lengths and angles are severe outliers. In particular, we retain for subsequent pruning and minimization all residues whose nearest neighbor alpha carbon distances are between 3.5 and 5.0 Å, and whose valence angles after concatenation and 10 steps of minimization are between 90° and 150°. This typically reduces the number of possible structures by 20–60%.

Relaxation

Each structure obtained by concatenation is minimized for 350 steps using the CHARMm potential. This relaxes many, though not all, of the unfavorable structures. The minimized structures are screened for internal energy, and all those having internal energies greater than 30% above the minimum are discarded. The retained structures are subjected to free energy evaluation, and ranked according to their free energies.

Treatment of crystal structures

Our side-chain conformational free energy search indicates that crystal structures with charged upward-pointing side chains were not at their free energy minimum. Such side chains are E4 in the HIV1 RT peptide and N5 in the SEV9 peptide. A similar problem is seen for R1 and K4 of the histone peptide in B27 that is based on a structure with a mixture of bound peptides. The side chains were adjusted to minimize the free energy of the crystal peptide, and to obtain a consistent data set for comparison.

Acknowledgments

We thank Dr. Robert Bruccoleri for providing the CONGEN program. This work was supported by grant AI30535-05 awarded to C.D. by the National Institute of Allergy and Infectious Diseases.

References

Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. 1983. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J Comput Chem 4*:187–217.

Bruccoleri RE, Novotny J. 1992. Antibody modeling using the conformational search program CONGEN. *Immunomethods 1*:96–106.

Caflisch A, Miranker A, Karplus M. 1993. Multiple copy simultaneous search and construction of ligands in binding sites: Application to inhibitors of HIV-1 aspartic proteinase. *J Med Chem 36*:2142–2164.

Caflisch A, Niederer P, Anliker M. 1992. Monte-Carlo minimization with thermalization for global optimization of polypeptide conformations in Cartesian coordinate space. *Proteins Struct Funct Genet 14*:102–109.

Cornette J, King B, Silverman M, DeLisi C. 1993. Graphical representations of class I MHC cleft. *J Mol Graphics 11*:174–179.

Falk K, Rotzschke O, Stevanovic S, Jung G, Rammensee HG. 1991. Allele-specific motifs revealed by sequencing of selfpeptides eluted from MHC molecules. *Nature 351*:290–296.

Fremont DH, Matsumura M, Stura EA, Peterson PA, Wilson IA. 1992. Crystal structures of two viral peptides in complex with murine MHC class I H-2k^b. *Science 257*:919–927.

Goodsell DS, Olson AJ. 1990. Automated docking of substrates to proteins by simulated annealing. *Proteins Struct Funct Genet 8*:195–202.

Gulukota K, Vajda S, DeLisi C. 1996. Peptide docking using dynamic programming. *J Comput Chem 17*:418–428.

Hart TN, Read NJ. 1992. A multiple-start Monte Carlo docking method. *Proteins Struct Funct Genet 13*:206–222.

Krystek S, Stouch T, Novotny J. 1993. Affinity and specificity of serine endopeptidase-protein inhibitor interactions. *J Mol Biol 234*:661–679.

Lee B, Richards FM. 1991. The interpretation of protein structures: Estimation of static accessibility. *Biopolymers 31*:993–1008.

Madden DR, Garboczi DN, Wiley DC. 1993. The antigenic identity of peptide-MHC complexes: A comparison of the conformations of five viral peptides presented by HLA-A2. *Cell 75*:693–708.

Madden DR, Gorga JC, Strominger JL, Wiley DC. 1991. The structure of HLA-B27 reveals nonamer self-peptides bound in an extended conformation. *Nature (Lond) 353*:321–325.

Miranker A, Karplus M. 1991. Functionality maps of binding sites: A multiple copy simultaneous search method. *Proteins Struct Funct Genet 11*:29–34.

Novotny J, Bruccoleri RE, Saul FA. 1989. On the attribution of binding energy in the antigen-antibody complexes McPC 603, D1.3 and HyHel-5. *Biochemistry 28*:4735–4749.

Rammensee HG, Friede T, Stevanovic S. 1995. MHC ligands and peptide motifs: First listing. *Immunogenetics 41*:178–228.

Rosenfeld R, Zheng Q, Vajda S, DeLisi C. 1993. Computing the structure of bound peptides: Application to antigen recognition by class I major histocompatibility complex receptors. *J Mol Biol 234*:515–521.

Rotstein SH, Murcko MA. 1993. GroupBuild: A fragment-based method for de novo drug design. *J Med Chem 36*(12):1700–1710.

Sezerman U, Vajda S, Cornette J, DeLisi C. 1993. Towards computational determination of peptide-receptor structure. *Protein Sci 2*:1827–1846.

Vajda S, Weng Z, Rosenfeld R, DeLisi C. 1994. Effect of conformational flexibility and solvation on receptor-ligand binding free energies. *Biochemistry 33*:13977–13988.

Vasmatzis G, Brower R, DeLisi C. 1994. Predicting immunoglobulin like hypervariable loops. *Biopolymers 34*:1669–1679.

Yue S. 1990. Distance-constrained molecular docking by simulated annealing. *Protein Eng 4*:177–184.

Zhang C, Vasmatzis G, Cornette JL, DeLisi C. 1996. Determination of atomic desolvation energies from the structures of crystallized proteins. In prep.

Zheng Q, Rosenfeld R, Vajda S, DeLisi C. 1993. Determining protein loop conformation using scaling-relaxation techniques. *J Comput Chem 14*: 556–565.