Effect of Conformational Flexibility and Solvation on Receptor−Ligand Binding Free Energies†

Sandor Vajda, Zhiping Weng, Rakefet Rosenfeld, and Charles DeLisi*

Department of Biomedical Engineering, Boston University, 44 Cummington Street, Boston, Massachusetts 02215

Received July 6, 1994; Revised Manuscript Received September 1, 1994®

ABSTRACT: A coherent framework is presented for determing the free energy change accompanying ligand binding to protein receptors. The most important new feature of the method is the contribution of the flexibility of the free ligand, and hence its conformational change on binding, to the free energy. Flexibility introduces two additional terms in the free energy difference: the internal energy difference between the ligand in the bound and free states and the backbone entropy loss. The former requires taking explicit account of the difference in solvation of the various forms of the free ligand. The solvation free energy change is estimated using an atomic solvation parameter model [Eisenberg & Mclachlan (1986) *Nature 319*, 199−203], with an improved parameter set. In order to evaluate the method, we applied it to three data sets for which increasingly general methods are required. The set to which the most restrictive theory can be applied consists of eight crystallized endopeptidase−protein inhibitor complexes which do not change conformation on binding and for which the major contribution to the solvation free energy is entropic. The results are in good agreement with the measured values and somewhat better than those previously reported in the literature. The second data set compares the relative binding free energies of biotin and its analogs for streptavidin. In this case the structures are also rigid, but solvation free energy must include both enthalpic and entropic components. We find that differential free energy predictions are approximately the same as those obtained by free energy perturbation techniques. The final application is an analysis of the measured stabilities of 13 different MHC receptor−peptide complexes. In this case we show that flexibility contributes 30−50% of the free energy change and find a correlation of 0.88 between our predicted free energies and peptide dissociation times.

An important goal at the interface of the cellular and molecular sciences is to develop a methodology that will permit a predictive understanding of the recognition of flexible ligands by receptors. By predictive understanding we mean ascertaining, without recourse to direct experimentation, the effect that changes in particular side chains, or other functional groups, have on the stability of complexes. Achieving this goal, even under limited sets of circumstances, has direct bearing on drug design strategies, as well as on a host of fundamental questions related to the nature of complexes that trigger cell activation.

In this paper we address the problem of macromolecular stability by developing computational tools for determining the binding free energies of known complexes in aqueous environments. The quantities of interest are the absolute and relative binding free energies, the former denoting the free energy difference between complexed and solubilized forms of two molecules and the latter denoting the difference between the binding free energies of two closely related complexes, such as ligands differing in a single chemical group.

Relative binding free energies are usually calculated by applying free energy perturbation (FEP) or thermodynamic integration (TI) methods (Mezei & Beveridge, 1966; van Gunsteren, 1988; Reynolds et al., 1992; Miyamoto & Kollman, 1992). These procedures are computationally demanding but effective when structures are known and relatively small. When structures are unknown *a priori*, such as in docking calculations, FEP or TI calculations would have to be performed repeatedly during the exploration of a large translational and rotational space for the complex of minimum free energy. Since a single free energy calculation by FEP or TI methods typically requires hours of supercomputer time even when structures are known, coupling them with docking (e.g., Goodsell & Olson, 1990; Leach & Kuntz, 1992; Caflish et al., 1992; Hart & Read, 1992; Cherfils & Janin, 1993; Stoddard & Koshland, 1993; Sezerman et al., 1993; Rosenfeld et al., 1993) or similarly complex conformational search procedures is beyond current computational capabilities (Wilson et al., 1991).

An alternative procedure, which is somewhat less rigorous but computationally feasible, is to use a two-step process, the first step generating some manageable number of low-energy conformations by using a standard energy function that omits hydration and the second step ranking the structures using an empirical free energy function (Novotny et al., 1989; Horton & Lewis, 1992; Wilson et al., 1991, 1993; Avbelj, 1992; Murphy et al., 1993; Smith & Honig, 1994).

In addition to the contributions typically included in molecular mechanics energy functions, an adequate free energy function must take account of solvation, as well as entropy associated with side-chain conformations, translational/rotational motion, and dilutional (cratic) effects (Novotny

† This work was supported by a grant from the NIH/NIAID (AI30535). The calculations on streptavidin were supported by a grant from the DOE (FG02-93ER61656) to S.V.

* To whom correspondence should be addressed.

® Abstract published in *Advance ACS Abstracts,* November 1, 1994.

0006-2960/94/0433-13977$04.50/0

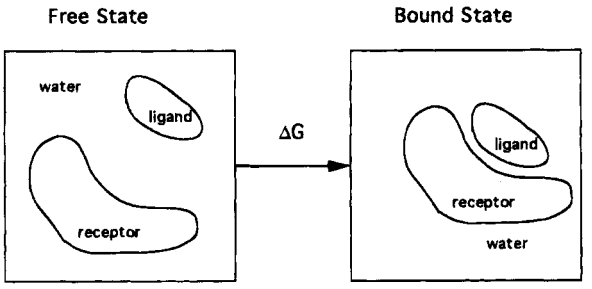

FIGURE 1: Formation of the receptor–ligand complex in the presence of water. All subsystems with energies included in the balance equations (eqs 1–3) are shown.

et al., 1989). Free energies obtained when these terms are included, using a solvation term proportional to the surface area (Hermann, 1972; Reynolds et al., 1974; Chothia, 1974; Novotny et al., 1989), and assuming rigid backbone geometries which are unaffected by reaction, compare well with observed values under appropriate conditions (Krystek et al., 1993). As we show below, however, the use of a solvation term that takes account of enthalpy as well as entropy is often important and sometimes crucial, since distinguishing side-chain type only by size underestimates the favorable effect of burying a hydrophobic group of the ligand in a hydrophobic pocket of the receptor. An equally serious requirement is allowance of backbone flexibility of free peptide ligands, especially when they are in the size range of small and intermediate size peptides.

We develop a methodology that begins with the method of Novotny et al. (1989) but relieve some restrictions of the original model by including energetic contributions to the solvation term and by estimating the contributions of free ligand flexibility to the free energy change. The method is tested against two sets of data for which the structures and reaction free energies are known. Both conform to the condition that the backbone structures remain unchanged by reaction. One consists of nine serine endopeptidase–inhibitor complexes; the other, streptavidin–biotin analog complexes.

The first set has been previously analyzed by Krystek et al. (1993), who approximated the change in solvation free energy by the change in solvent-exposed surface. In this case, our calculations and theirs yield similar results. In the second case, predictions of relative free energies based on our modified solvation funtion are within 6.8% of observed values for both analogs and within 2.4% of predictions based on FEP methods, whereas a free energy function that does not allow for the enthalpic contribution to solvation is in error by over 65% for the polar analog. Finally, we compare predictions to data on the stability of 13 different flexible nonameric peptides interacting with a class I major histocompatibility complex (MHC) receptor and find a strong correlation between our calculated free energies and experimental estimates of the stabilities of these complexes.

METHODS

We are interested in the free energy change accompanying the formation of a receptor–ligand complex in the presence of water (Figure 1). The subscripts r, l, and w label receptor, ligand, and water, respectively, and the superscripts f and b denote quantities prior and subsequent to complex formation (i.e., free and bound states). The energy in the free state is given by

$$E^{f} = E^{f}_{r} + E^{f}_{l} + E^{f}_{w} + E^{f}_{r-w} + E^{f}_{l-w} \quad (1)$$

where the last two terms refer to receptor–water and ligand–water interactions. After complex formation

$$E^{b} = E^{b}_{r} + E^{b}_{l} + E^{b}_{w} + E^{b}_{r-l} + E^{b}_{(rl)-w} \quad (2)$$

the last two terms referring to the interaction energy between receptor and ligand and the interaction energy between water and the receptor–ligand complex. All terms involving water refer to the local environment of the macromolecules; i.e., interactions with and between water molecules that are not in the vicinity of macromolecular surfaces are assumed unchanged by complex formation. The energies of the isolated molecules include all bonded and nonbonded interactions.

In order to make the connection with previous research (Novotny et al., 1989; Horton & Lewis, 1992), we begin by confining the development to systems in which the ligand and receptor have known structures that are unchanged by binding. After developing the formalism and testing it within this restricted domain, we discuss and analyze the more general and more difficult problem in which the ligand structure changes on binding.

Rigid Structures

The Energy Contribution to Free Energy. The quantity of interest, the energy difference between the bound and free states, is obtained by subtracting eq 1 from eq 2:

$$\Delta E = \Delta E_{r} + \Delta E_{l} + E^{b}_{r-l} + [E^{b}_{(rl)-w} - E^{f}_{r-w} - E^{f}_{l-w} + \Delta E_{w}] \quad (3)$$

where $\Delta E_{w} \equiv E^{b}_{w} - E^{f}_{w}$ takes into account possible changes in the self-energy of water, $\Delta E_{r} \equiv E^{b}_{r} - E^{f}_{r}$, and $\Delta E_{l} \equiv E^{b}_{l} - E^{f}_{l}$. The bracketed term on the right, the hydration or solvation energy, is the energy difference between the solvated ligand–receptor complex and the sum of the solvated receptor and solvated ligand when they are widely separated:

$$\Delta E_{h} \equiv E^{b}_{(rl)-w} - E^{f}_{r-w} - E^{f}_{l-w} + \Delta E_{w} \quad (4)$$

Each of the terms in eq 4 is a sum over reaction energies of all atoms in the complex, and each interaction energy is written as the sum of electrostatic and van der Waals contributions. For those systems in which the structures remain unchanged by the reaction, the first two terms on the right in eq 3 are set equal to zero.

A fundamental and widely used assumption is that the protein–protein and protein–solvent interfaces are well packed to the extent that the van der Waals component of Δe is relatively small (Adamson, 1976; Novotny et al., 1989; Nicholls et al., 1991; Krystek et al., 1993). In that case we replace ΔE_{h} and E^{b}_{r-l} by their electrostatic components, ΔE^{el}_{h} and E^{el}_{r-l}, and eq 3 becomes

$$\Delta E = E^{el}_{r-l} + \Delta E^{el}_{h} \quad (5)$$

The Entropy Contribution. In order to obtain an expression for the total free energy difference, eq 5 must be

augmented by the addition of entropic terms. These include ΔS_{tr}, the rotational and translational entropy change; ΔS_{cr}, the cratic (dilutional) entropy change; ΔS_{sc}, the conformational entropy change of the side chains that become buried upon complexation; and ΔS_h, the hydrophobic entropy change (the cost of creating a cavity in the solvent). The full free energy expression is then

$$\Delta G = E^{el}_{r-l} + \Delta G_h - T\Delta S_{sc} - T\Delta S_{tr} - T\Delta S_{cr} \quad (6)$$

where $\Delta G_h = \Delta E^{el}_h - T\Delta S_h$ is the hydrophobic free energy.

The translational/rotational and cratic contributions to the entropy are estimated as ≈7 and ≈2 kcal/mol, respectively (Erickson, 1989; Novotny et al., 1989) and are in any case generally taken to be an additive constant, to a good first approximation independent of the nature of the complex. The hydrophobic and conformational entropy terms, on the other hand, are highly sensitive to the chemical nature of the interacting molecules. These are treated as follows.

(A) Hydrophobic Entropy. The calculation of hydrophobic free energies is based on the procedure of Eisenberg and McLachlan (1986) for calculating the free energy of transferring a molecule from octanol to water. Briefly, an average solvent-exposed surface area is calculated for each of the five types of atoms—charged (N^+ or O^-), polar uncharged (N/O or S), and apolar (C)—in each of the 20 side-chain types (Shrake & Rupley, 1973). The octanol to water transfer free energy for each (terminally blocked) amino acid residue is then written as a linear combination of these average exposed surface areas.

The weights or atomic solvation parameters (ASP's), obtained by a least squares fit to the transfer free energy data based on observed water/octanol partition coefficients (Eisenberg & McLachlan, 1986), were recently recalculated by Pickett and Sternberg (1993) using a reinterpretation of experimental hydrophobicity values by Sharp et al. (1991). Estimates of the hydrophobicity of Cys and His based on these parameters have large residual errors. Cys is probably in error because a reliable value for −SH is unavailable (Eisenberg & McLachlan, 1986); His is in error because its ionization state is uncertain. We therefore repeated the fitting of the model to the transfer free energy data corrected by Sharp et al. (1991) by excluding these two residue types. Using the notation of Eisenberg and McLachlan (1986), in cal mol^{-1} $Å^{-2}$ units, the ASP values are as follows: $\Delta\sigma(C) = 31 \pm 2$, $\Delta\sigma(N/O) = -1 \pm 4$, $\Delta\sigma(O^-) = -15 \pm 12$, $\Delta\sigma(N^+) = -38 \pm 11$, and $\Delta\sigma(S) = 10 \pm 16$.

The difference in solvation free energies between polar and apolar side chains calculated with these coefficients is greater than what is obtained using classical methods that assume proportionality to surface area (Hermann, 1972; Reynolds et al., 1974; Chothia, 1974; Novotny et al., 1989). This is due to the enthalpic contribution to polar side chains in water, which is included implicitly in the ASP model but not in the surface area model. For example, using the above solvation parameters and the average atomic areas of Shrake and Rupley (1973), the calculated free energy of transferring Ile from octanol to water is 5.1 kcal/mol, whereas the free energy of transferring Asp is 0.98 kcal/mol. If 0.025 × surface area is used (Reynolds et al., 1974; Novotny et al., 1989), the results are 4.0 and 2.9 kcal/mol, respectively; i.e., the Asp transfer is relatively less favorable than it should be because of the omission of the enthalpic contribution.

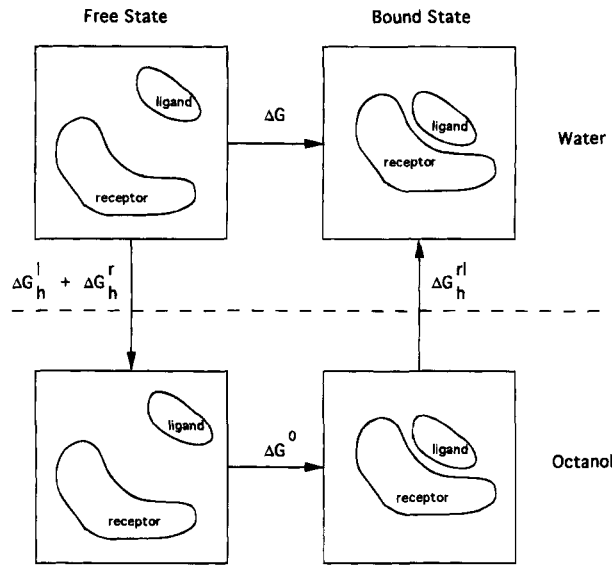

FIGURE 2: Thermodynamic cycle used in the calculation of the binding free energy ΔG by eq 7. The binding free energy in octanol, $\Delta G°$, is assumed to include all free energy terms except for the hydrophobic free energy $\Delta G_h = \Delta G^{rl}_h - \Delta G^l_h - \Delta G^r_h$.

With the weights determined, the transfer free energy of any molecule can be obtained by determining the solvent-accessible areas of each atom type, multiplying by the weights, and summing over the atoms. For the results presented below, the areas are obtained using the Lee–Richards algorithm (1978), except for streptavidin whose binding orifice cannot accommodate the 1.4-Å sphere used to represent water. In that case we used the method of Connolly (1983).

With the transfer free energies for each amino acid residue determined, the transfer free energies of the complex (ΔG^{rl}_h), the receptor (ΔG^r_h), and the ligand (ΔG^l_h) are easily obtained, and the hydrophobic contribution to the free energy change is given by

$$\Delta G_h = \Delta G^{rl}_h - \Delta G^l_h - \Delta G^r_h \quad (7)$$

Notice that the hydrophobic free energy is approximated by the free energy of transfer from octanol to water, rather than from vacuum to water, since the van der Waals terms are assumed to cancel. Thus our free energy evaluation model is based on the thermodynamic cycle shown in Figure 2.

(B) Conformational Entropy: The Free State. The entropy S^f_j of the free state for a residue of type j is evaluated using the classical relationship

$$S^f_j = -R\sum_i p_{ij} ln(p_{ij}) \quad (8)$$

where the p_{ij} values, the probability of side-chain type j being in conformational state i, were recently estimated by Pickett and Sternberg (1993) from the observed distributions of exposed side chains in proteins with known X-ray structures. Pickett and Sternberg consider a side chain exposed when its accessibility is more than 60% of the standard side-chain surface area (Lee & Richards, 1978). The substitution of the estimated probability values into eq 8 yields values for the entropic free energy term TS^f ranging from 0 (for Ala, Gly, and Pro) to 2.1 kcal/mol (for Glu). These estimates are comparable to, but generally smaller than, those obtained

using three equal states per rotatable bond (Novotny et al., 1989). Entropy changes are calculated for the side chains of both the receptor and the ligand.

(*C*) *Conformational Entropy: The Bound State.* Ligand binding will bury, to various degrees, side chains of both the ligand and the receptor, restricting motion about otherwise rotatable bonds. In order to obtain entropy change estimates for partially buried side chains, we set $\Delta S = \Delta S_{max} = S^f$ if the change ΔA_t in the total solvent-accessible surface area of the side chain is more than 60% of the standard side-chain surface area A_t^*. Otherwise the entropy loss is scaled according to $\Delta S = \alpha \Delta S_{max}$, where $\alpha = \Delta A_t/(0.6A_t^*)$. Thus, the loss of side-chain conformational entropy is assumed to be proportional to the loss in solvent-accessible surface, with the proportionality constant chosen so that full entropy loss occurs when 60% of the standard solvent-accessible surface becomes buried.

To summarize, we estimate the binding free energy between a rigid ligand and a rigid receptor by

$$\Delta G = \text{const} + E^{el}_{r-l} + \Delta G_h - T\Delta S_{sc} \quad (9)$$

where the constant term includes cratic, rotational, and translational contributions to entropy, E^{el}_{r-l} is the electrostatic component of the receptor−ligand interaction energy, ΔG_h is the solvation contribution (eq 7) to the binding free energy, and $T\Delta S_{sc}$ is the conformational entropy loss of side chains in the receptor and the ligand. Note that the solvation energy ΔE_h defined by eq 4 is not calculated explicitly but is included in the empirically estimated hydrophobic free energy, ΔG_h.

When X-ray structures are analyzed, the free energy contributions to eq 9 are calculated after 200 steps of ABNR (adopted basis Newton−Raphson) minimization of the complex using Version 22 of the CHARMm potential (Molecular Simulations, Inc.) with a distance-dependent dielectric coefficient $\epsilon = 4r$ and a nonbonded cutoff of 17 Å. Mass-weighted harmonic constraints with a force constant of 20 kcal/mol^{-1} Å^{-2} are applied to all atoms. This constrained minimization eliminates steric overlaps and severe deviations from standard polypeptide geometry, with only minor deviation from the unrefined structure.

Flexible Ligands

Although a number of systems conform to the assumption of relatively rigid backbone geometry in which the reactive units do not change conformation on complexation, the assumption is clearly restrictive. In particular, it does not hold when the ligand is a flexible peptide. In that case the free ligand has no dominant energy state and the appropriate quantity to consider is its energy averaged over a representative sample of conformations, while the bound ligand has a dominant state induced by interaction with the receptor. This more general circumstance introduces two additional terms in the free energy difference: the internal energy difference between the ligand in the bound and free states and the backbone entropy difference.

The Internal Energy Difference. Because the free ligand is flexible, neither its backbone entropy change nor its internal energy change can be set equal to zero, nor can its hydrophobic transfer free energy be assigned a value on the basis of a single conformation. The ligand's internal energy difference must therefore be averaged over a large population of conformations, as must its solvation free energy difference. Consequently, eq 9 is replaced by

$$\Delta G = \text{const} + E^{el}_{r-l} + \Delta G^{rl}_h - \Delta G^{r}_h - \langle \Delta G^{l}_h \rangle + \langle \Delta E_l \rangle - T\Delta S_{sc} - T\Delta S_{bb} \quad (10)$$

where ΔS_{bb} is the backbone entropy change, and

$$\langle \Delta G^{l}_h - \Delta E_l \rangle = \sum_i (\Delta G^{l}_h - \Delta E_l)_i \exp[-(\Delta G^{l}_h - \Delta E_l)_i/RT]/Q \quad (11)$$

with Q being the partition function:

$$Q = \sum_i \exp[-(\Delta G^{l}_h - \Delta E_l)_i/RT] \quad (12)$$

As in the case of a rigid ligand, we assume that solute atoms are well packed by other solute or solvent atoms, so van der Waals packing is nearly the same for all conformations, irrespective of whether the peptide is bound or free. Thus the van der Waals energy for any free conformation is assumed to cancel the van der Waals energy for the bound conformation, and eq 11 becomes

$$\langle \Delta G^{l}_h - \Delta \tilde{E}_l \rangle = \sum_i (\Delta G^{l}_h - \Delta \tilde{E}_l)_i \exp[-(\Delta G^{l}_h - \Delta \tilde{E}_l)_i/RT]/\tilde{Q} \quad (13)$$

where the tilde denotes the soft (i.e, non van der Waals) components of the energy function and the partition function $\tilde{Q}$ is calculated by using $(\Delta \tilde{E}_l)_i$ in eq 12. In order to avoid spuriously high values for bond length and bond angle terms in the conformational energy of the bound ligand, the peptide−receptor energy is first minimized for 100 steps with the receptor atoms constrained.

The averages in eq 13 were calculated over a sample of 9000 conformations which was generated by selecting the backbone dihedral angles from an empirical distribution based on X-ray structures of proteins and by selecting side-chain torsional angles from a uniform distribution between −180° and +180°. Each peptide conformation was minimized for 1000 steps without constraints using the CHARMm potential (Brooks et al., 1983).

Although the hydrophobic free energy $(\Delta G_h)_i$ of the ligand calculated by the ASP model (Eisenberg & MacLachlan, 1986) takes account of the solvation free energy, the absence of explicit solvent inevitably biases the sample. This bias shows through the disproportionately favorable energies of hairpin-like structures having oppositely charged end groups in proximity. Such charged groups are ordinarily solvated, thereby screening electrostatic interactions at relatively small distances and also preventing, through solvation shell intervention, charges from approaching each other too closely. In this initial effort, we prevent close approach somewhat arbitrarily by discarding structures having end charges closer than 5 Å—a distance that would approximate the sum of van der Waals radii of the end groups with an intervening water molecule.

The final distributions typically include approximately 8000 structures. The distribution of energy states is such that a relatively small number (usually less than 30) of structures near the minimum dominate the average. In order

Table 1: Experimental and Calculated Binding Free Energies for Nine Serine Endopeptidase−Protein Inhibitor Complexes Studied by Krystek et al. (1993)[a]

PDP code[b]	E^{el}_{r-l}	ΔG_h	$-T\Delta S_{sc}$	ΔG_{cal}	ΔG_{exp}	ΔE_h
4SGB	−8.47 (−11.2)	−23.43 (−35.1)	9.40 (21.6)	−13.50 (−13.7)	−11.7	11.7
3SGB	−12.47 (−15.6)	−21.20 (−35.1)	14.19 (25.6)	−10.58 (−14.1)	−12.7	13.9
1CHO	−17.58 (−15.6)	−25.42 (−40.3)	19.48 (30.6)	−14.52 (−14.3)	−14.4	14.9
2SNI	−15.50 (−18.6)	−30.58 (−44.7)	21.46 (35.4)	−15.62 (−16.9)	−15.8	14.1
2SEC	−16.03 (−22.0)	−26.14 (−40.5)	17.53 (33.6)	−15.64 (−17.9)	−14.0	14.1
1TEC	−14.90 (−18.6)	−27.44 (−41.7)	18.82 (33.6)	−14.52 (−15.7)	−14.0	14.3
2PTC	−23.94 (−30.0)	−19.39 (−37.5)	17.13 (36.0)	−17.20 (−21.7)	−18.1	18.1
2TGP	−22.71 (−29.2)	−20.19 (−37.6)	16.81 (36.0)	−17.09 (−23.4)	−7.9	17.4
2KAI	−22.14 (−53.0)	−21.49 (−38.5)	18.65 (37.2)	−15.98 (−43.4)	−12.5	17.0

[a] Energies are in kcal/mol. Values calculated by Krystek et al. (1993) are shown in parentheses. The solvation energy is calculated by $\Delta E_h = \Delta G_h - 0.025\Delta A_t$, where A_t denotes the total contact area (see Discussion). [b] 4SGB, proteinase B with potato chymotrypsin inhibitor 1; 3SGB, proteinase B with turkey ovomucoid inhibitor third domain; 1CHO, α-chymotrypsin with turkey ovomucoid inhibitor third domain; 2SNI, subtilisin novo with barley seed chymotrypsin inhibitor 2; 2SEC, subtilisin Carlsberg with eglin-*c*; 1TEC, thermitase with eglin-*c*; 2PTC, trypsin with bovine pancreatic trypsin inhibitor; 2TGP, trypsinogen with bovine pancreatic trypsin inhibitor; 2KAI, kallikrein with bovine pancreatic trypsin inhibitor.

to reduce the possibility of error due to sparse statistics near the minimum, for each peptide we choose the three lowest energy conformations found and for each of these generate 1000 additional structures by slightly perturbing (±20°) all dihedrals (backbone and side chain) in the structure, followed by 600 steps of energy minimization. After discarding structures having end charges closer than 5 Å, each of the starting structures yielded at least 300 conformations retained, thus almost 1200 conformations in the three runs. The $(\Delta G^l_h - \Delta \tilde{E}_l)$ values of these structures follow a relatively smooth distribution, with at least 100 of them significantly contributing to the ensemble average calculated by eq 13.

The Conformational Entropy. The second additional effect that enters the free energy when ligands are flexible is the loss of backbone entropy. As in the case of side chains, the backbone entropy loss can be estimated computationally with varying degrees of sophistication. The simplest method involves an empirical approach, similar to the one used for calculating the loss of side-chain conformational entropy (Pickett & Sternberg, 1993).

If the peptide backbone loses its entire conformational entropy upon binding to the receptor, the backbone conformational entropy loss ΔS_{bb} can be obtained from an analog of eq 8. In particular, we use a recently developed method in which 16 regions of the (ϕ,ψ) map are defined as backbone conformational states, and the probabilities of these states are predicted by using forward and backward Markov models (Vajda, 1993). Although the transitional probabilities used in these models were estimated from a data base of high-resolution X-ray structures of globular proteins, they can also be used to determine conformational probabilities for the residues of short linear peptides in solution, and the predictions are in good agreement with NMR data (Vajda, 1993). Using the probabilities estimated for every residue of the given sequence, the loss of backbone entropy is calculated by eq 8.

RESULTS

The method was applied to three different systems. (1) Crystallized endopeptidase−inhibitor complexes are the best characterized experimentally and satisfy the rigid body, no conformational change requirement. These have also been analyzed previously by Novotny and his colleagues (Krystek et al., 1993). (2) The biotin−streptavidin system is also well characterized experimentally. In addition, it allows us to compare our method with FEP techniques (Miyamoto & Kollman, 1993) and to test the effect of our modified approach to treating solvation. (3) The stability of class I MHC complexed with antigenic peptides has been characterized in terms of dissociation times (Parker et al., 1992a). Binding free energy calculations for this system require allowance for peptide flexibility.

Serine Endopeptidase−Inhibitor Complexes. Good agreement between calculated and observed free energy values was obtained for nine crystallized serine endopeptidase−inhibitor complexes using either the method of Krystek et al. (1993) or our modification (Table 1). The observed and calculated values are within 2 kcal/mol for seven of the nine complexes, the outliers being the trypsinogen−BPTI complex (2TGP) and the kallikrein−BPTI complex (2KAI).

The 2TGP complex is included in the table for the sake of completeness. It does not meet the assumption of rigid molecular association, since trypsinogen changes conformation when it complexes with BPTI (Huber & Bode, 1978). The sources of overestimating the free energy of 2KAI by 3.4 kcal/mol will be discussed below. With these two complexes omitted, the RMSD of our free energy estimates is 1.3 kcal/mol, and that of Krystek et al. (1993) is 2.4 kcal/mol.

Binding of Biotin and Its Analogs to Streptavidin. The structure of streptavidin, with and without biotin, has been determined by X-ray crystallography (Weber et al., 1989; Hendrickson et al., 1989). The binding free energy of biotin for streptavidin (Figure 3) is −18.3 kcal/mol, among the highest for any known noncovalently associated structure (Weber et al., 1992). Binding free energies are also known for a number of biotin analogs, including iminobiotin and thiobiotin in which biotin's ureido oxygen is replaced respectively by an NH group and a sulfur (Green, 1966, 1975, 1990). The ureido oxygen participates in three hydrogen bonds with the side chains of residues Asn-23, Ser-27, and Tyr-43 (Figure 3).

Models of biotin and its analogs were constructed using the geometric and force field parameters of the QUANTA CHARMm program (Molecular Simulations, Inc.). We present only relative binding free energies $\Delta\Delta G$, i.e., the difference in the free energy between biotin and its analogs. This eliminates the need to calculate entropic contributions. Indeed, the single atom substitutions leave the biotin entropy essentially invariant. Furthermore, the substitutions do not change the overall shape of the ligand, and the changes in partial charges are restricted to the atoms involved and their

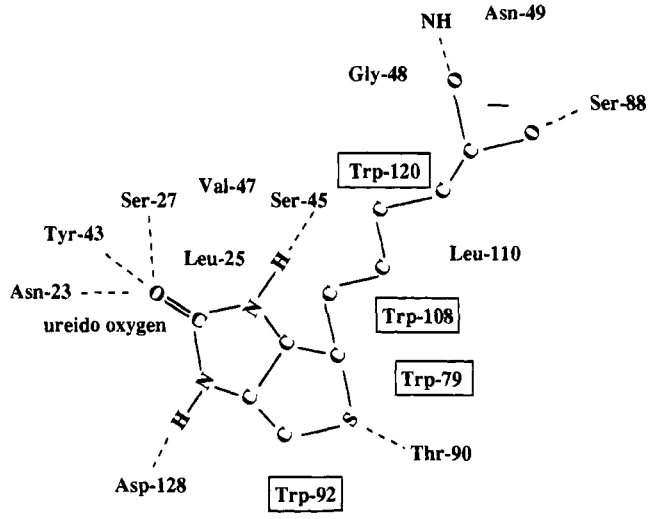

FIGURE 3: Schematic representation of biotin and the surrounding residues of streptavidin. Hydrogen bonds between biotin and streptavidin atoms are shown in dashed lines. The ureido oxygen participates in three hydrogen bonds with the side chains of residues Asn-23, Ser-27, and Tyr-43. The streptavidin-bound biotin is almost completely buried; only the valeryl carboxyl oxygens (on the upper right side of the figure) and one NH group are partially available to solvent. Substituting the ureido oxygen by an NH group or by an S atom yields iminobiotin and thiobiotin, respectively.

Table 2: Relative Binding Free Energies of Biotin Analogs to Streptavidin[a]

compound	E^{el}_{r-l}	ΔE^{el}_{r-l}	$\Delta\Delta G_h$	$\Delta\Delta G_{cal}$	$\Delta\Delta G_{obs}$	$\Delta\Delta G_{FEP}$
biotin	−8.61					
iminobiotin	−4.64	3.97	2.62	6.59	6.10	6.50
thiobiotin	−5.46	3.15	0.16	3.31	3.50	3.20

[a] Energies are in kcal/mol. $\Delta\Delta G_{cal}$ and $\Delta\Delta G_{obs}$ denote calculated and observed values, respectively, of the relative binding free energy, defined as the difference between the binding free energy for the biotin analog and that for biotin. $\Delta\Delta G_{obs}$ is based on avidin data (Green, 1966). $\Delta\Delta G_{FEP}$ is the relative binding free energy calculated by a free energy perturbation (FEP) method (Miyamoto & Kollman, 1993).

nearest neighbors. In particular, five hydrogen bonds and the strong hydrophobic interaction of biotin and its analogs with four tryptophan side chains remain unaffected by the changes (Figure 3). Therefore, it is reasonable to assume that biotin, iminobiotin, and thiobiotin occupy essentially the same location in the binding site, and hence the side-chain entropy loss terms cancel in $\Delta\Delta G$.

The results (Table 2) indicate that the calculated relative binding free energies are about as accurate as the values recently obtained by free energy perturbation methods (Miyamoto & Kollman, 1993). The primary advantage of the free energy perturbation in this case is its ability to provide the absolute free energy of biotin binding. However, the calculation of relative free energies by perturbation methods requires 2−3 orders of magnitude more CPU time than by empirical methods.

Unlike the predictions on the serine endopeptidase−inhibitor system in which the purely entropic hydrophobicity term is adequate (Krystek et al., 1993), the results for this system illustrate the importance of taking account of the enthalpic difference between transferring an apolar (in this case S) as opposed to a charged (NH) group from water to a largely apolar environment (the streptavidin pocket). Of the observed 3.3 kcal/mol free energy difference between the two analogs (Table 2), almost 2.5 kcal/mol is contributed by the difference in solvation free energy between the ureido oxygen and the charged NH group in iminobiotin on the one hand and the same oxygen and uncharged S in thiobiotin on the other. Since the surface areas are in both cases roughly the same, this contribution would be lost if a purely entropic term were used.

Binding of Peptide Antigens to MHC Class I Receptors. The class I MHC molecule is a heterotrimer consisting of a 44-kDa membrane-spanning chain, noncovalently associated with a 12-kDa β_2-microglobulin (β_2-m) chain and an antigenic peptide. Under certain experimental conditions the characteristic dissociation time of β_2-m can be used as a measure of the stability of the bound peptide (Parker et al., 1992a,b).

Equation 10 was used to calculate the binding free energies of five viral peptides for the HLA-A2 class I receptor. The crystal structure for each has been solved by Madden et al. (1993), and for four of them the β_2-m dissociation time has been measured (Parker et al., 1994). The results of the calculation (Table 3) indicate that the flexibility of the free peptide and its consequent stabilization on binding make substantial contributions to both the energy and entropy portions, typically reducing the overall free energy to half the value of the total free energy change it would have had were the structures rigid.

Additional stability data studied here are for nine peptides obtained from the influenza A virus matrix protein (Flu A) peptide by mutating each of the nine positions into Ala one by one (Parker et al., 1992b). In this case the crystal structures are not available, so we first had to calculate the bound structures of the mutant peptides. This was done by assuming that the backbone does not change and replacing the appropriate side chain by a methyl group. Since the backbone conformation uniquely defines the positions of the C_β atoms, the side-chain substitutions can be carried out without a conformational search. The binding free energies are calculated as for the X-ray structures (Table 4), although the substitution of a side chain by Ala may affect the backbone of the bound peptide, and hence the structures are likely to be somewhat less accurate. Nevertheless, a plot of binding free energy against dissociation time (Figure 4) for the pooled data set of 13 peptides (i.e., the four peptides from Table 3 and the nine Ala-substituted peptides in Table 4) shows a correlation that is well within experimental error.

DISCUSSION

We have presented and applied a coherent framework for calculating the free energy change accompanying the binding of rigid or flexible ligands to protein receptors. The empirical free energy function includes a realistic account of side-chain rotamer distributions, takes account of enthalpic as well as entropic contributions to the solvation free energy, and allows for flexibility of the ligand. Because the free energy function is empirical, the evaluation is rapid enough to allow its use in conjunction with docking calculations in which the bound conformation of the ligand is not known *a priori*. The modified treatment of side-chain entropy introduces a relatively minor change but correlates well with what appears in the literature. The modified treatment of solvation leads to substantive changes in predictions, as does ligand flexibility.

Table 3: Calculated Binding Free Energies and Observed Dissociation Times for HLA-A2 Complexed with Antigenic Peptides of Known X-ray Structure (Madden et al., 1993)[a]

peptide[b]	E^{el}_{r-l}	ΔG^{rl}_h	ΔG^{r}_h	$-T\Delta S_{sc}$	$-T\Delta S_{bb}$	$\tilde{E}^{b}_l$	$\langle \Delta G^{l}_h + \tilde{E}^{f}_l \rangle$	ΔG	half-life
HIV-1 RT	−37.53	109.97	113.64	19.58	3.78	9.61	12.39	−11.62	190
HTLV-1 Tax	−31.72	122.39	124.03	17.39	4.95	14.89	30.87	−18.00	6400
Flu A matrix	−24.42	113.64	118.30	16.75	5.38	3.02	18.84	−13.77	1000
HIV-1 gp120	−32.96	113.16	−117.44	21.41	7.90	−1.03	2.99	−5.97	N/A
Hep B capsid	−32.45	107.12	113.29	18.96	7.50	24.68	36.26	−14.74	1500

[a] All energies are in kcal/mol. The flexible contribution $\langle \Delta G^{l}_h - \Delta \tilde{E}_l \rangle$ defined by eq 13 is calculated as $\langle \Delta G^{l}_h + \tilde{E}^{f}_l \rangle - \tilde{E}^{b}_l$. ΔG denotes the calculated binding free energy. Half-lifes are in minutes (Parker et al., 1994). [b] The sequences of the peptides (using single-letter amino acid codes) are as follows: HIV-1 RT, ILKEPVHGV; HTLV-1 Tax, LLFGYPVYV; Flu A matrix, GILGFVFTL; HIV-1 gp120, TLTSCNTSV; Hep B capsid, FLPSDFFPSV.

Table 4: Calculated Binding Free Energies and Observed Dissociation Times for HLA-A2 Complexed with Ala-Substituted Analogs of the Influenza A Matrix Peptide[a]

position[b]	E^{el}_{r-l}	ΔG^{rl}_h	ΔG^{r}_h	$-T\Delta S_{sc}$	$-T\Delta S_{bb}$	$\tilde{E}^{b}_l$	$\langle \Delta G^{l}_h + \tilde{E}^{f}_l \rangle$	ΔG	half-life
G1	−24.38	113.64	118.32	16.76	5.05	3.86	19.34	−13.73	1800
I2	−24.40	113.68	118.30	15.84	5.32	2.48	10.81	−7.19	89
L3	−24.42	113.81	118.30	15.90	5.15	0.76	13.35	−11.45	240
G4	−24.42	114.41	118.31	16.83	5.71	3.46	17.16	−10.48	610
F5	−24.42	114.64	118.30	15.55	5.29	2.22	13.50	−9.52	220
V6	−24.42	113.24	118.30	16.23	4.78	2.61	16.02	−12.88	220
F7	−24.42	113.87	118.30	15.85	5.41	1.20	9.32	−6.81	97
T8	−22.61	113.81	118.31	14.98	4.96	2.34	15.81	−11.64	1000
L9	−24.41	113.80	118.30	15.91	5.49	2.68	15.85	−11.67	560

[a] Energies are in kcal/mol and half-lifes are in minutes (Parker et al., 1994). [b] Standard single-letter code of the amino acid substituted by Ala at the given position of the Flu A matrix peptide.

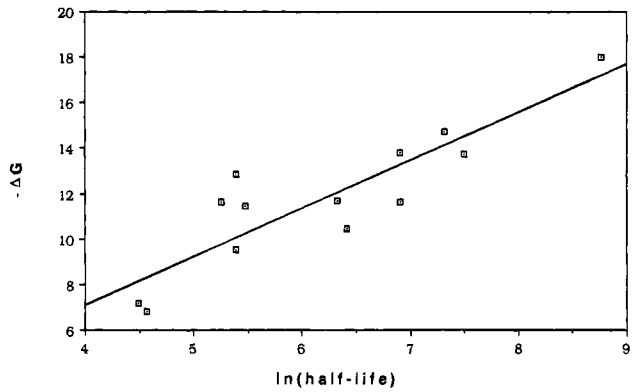

FIGURE 4: Calculated binding free energies and observed β_2-microglobulin dissociation times (Parker et al., 1994) for complexes of HLA-A2 with four antigenic peptides studied by Madden et al. (1993) and nine Ala-substituted analogs of the influenza A matrix peptide.

Solvation. The use of an expression of the form $\Delta G_h = \gamma \Delta A_t$ (where ΔA_t is the decrease of the total solvent-accessible surface area of the protein and $\gamma = 0.025$ kcal mol^{-1} Å^{-2}) to evaluate the solvation contribution to the total binding free energy change appears to be less effective than the atomic solvation parameter model of Eisenberg and McLachlan (1986). The latter represents ΔG_h as a linear combination of the changes in surface areas of five different atomic groups and implicitly includes enthalpic contributions since the parameters are based on measured transfer free energies. The term $\gamma \Delta A_t$, on the other hand, reflects predominantly the changes in the solvent entropy at the solvent−solute interface (Novotny et al., 1989). The relationship (Chothia, 1974) between accessible surface area and solvation free energy is in agreement with the clathrate model of solvation (Kauzman, 1959): solvating a macromolecule reduces the translational and rotational degrees of freedom of nearby water molecules, thereby reducing entropy, an effect that depends only on the solvent-accessible area and not on the local polar/apolar character of the surface. Polar and charged chains, however, interact with polar water, and hence the solvation energy ΔE^{el}_h can compensate for the entropy loss, resulting in a small positive solvation free energy change for polar and a negative free energy change for charged side chains, which is in agreement with the ASP model (Eisenberg & McLachlan, 1986).

If $\gamma \Delta A_t$ approximates the solvation entropy rather than the free energy, then we can estimate the electrostatic component of the solvation energy by $\Delta E^{el}_h = \Delta G_h - \gamma \Delta A_t$; i.e., the solvation energy ΔE^{el}_h is obtained as the difference between the solvation free energy calculated from the ASP model and the solvation free energy using the total area model. As shown in Table 1, the resulting ΔE^{el}_h value is always positive. Indeed, $\Delta E^{el}_h = E^{el}_{(rl)-w} - E^{el}_{r-w} - E^{el}_{l-w}$, where the terms on the right-hand side denote the complex−water, receptor−water, and ligand−water interaction energies. The sum $E^{rel}_{r-w} + E^{el}_{l-w}$ is always more negative than $E^{el}_{(rl)-w}$, because additional charges interact with polarized solvent, and hence the solvation energy ΔE^{el}_h is always positive.

Since both the electrostatic and solvation energies depend primarily on surface charges, we expected some correlation between the two quantities. For example, a simple picture would have the solvation energy proportional to the number of charged groups on the two surfaces, which become buried after complex formation. With perfectly complementary opposite charges on opposing surfaces, the electrostatic energy would also be proportional to the number of charged groups, and the correlation between the two would be very strong. The correlation would be diminished by deviations from perfect complementarity or by variation in subsurface charges from structure to structure, which could contribute to variations in the electrostatic energy but not to solvation. Our results, however (Figure 5), indicate that complemen-

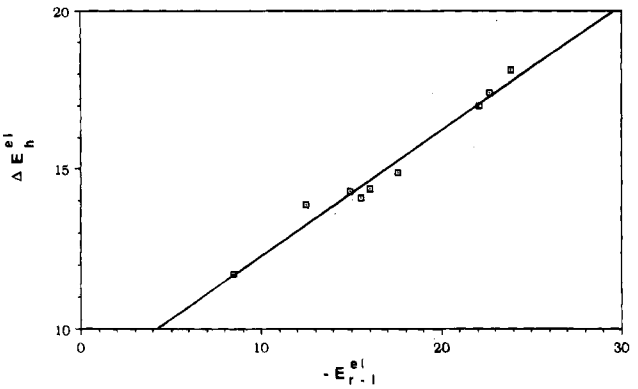

FIGURE 5: Correlation between the electrostatic interaction energy E^{el}_{r-l} and the hydrophobic energy ΔE^{el}_h for the nine serine endopeptidase−protein inhibitor complexes listed in Table 1.

tarity is very strong, the correlation coefficient being 0.97 for the nine endopeptidase−protein inhibitor complexes. In retrospect this is not entirely surprising since such complementarity minimizes the electrostatic part of the free energy and does so with only minor cost in terms of other contributions to the free energy.

Additional insight into the significance of this result is provided by the fact that neither $\gamma\Delta A_t$ nor ΔG_h alone correlate with E^{el}_{r-l}, the correlation coefficients being 0.10 and −0.44, respectively. The almost perfect, linear relation between E^{el}_{r-l} and ΔE^{el}_h lends further support to the assumption that $\gamma\Delta A_t$ estimates only the solvation entropy and not the entire solvation free energy.

A slight digression at this point might provide some perspective on the treatment of electrostatic contributions to the solvation of charged side chains. One approach would be to use an area term for the solvation free energy, supplemented with an electrostatic term calculated using some form of the Poisson−Boltzmann equation. In principle this can be done accurately for relatively simple systems, e.g., when geometries are unchanged by binding. For flexible systems, or as part of a general docking algorithm, such an approach would require repeated solutions to the Poisson−Boltzmann equation for very complex geometries, and it is therefore not presently viable.

The alternative approach used here relies on a modified ASP model for treating solvation, with an additional explicit potential term for electrostatic interactions between and within solute molecules. This appears to be reasonably effective, as the above correlation—as well as our comparisons with experiment—suggests, and is also in accord with results of Smith and Honig (1994), who found that, in a search for protein loop conformations, the two methods were comparable.

The correlation shown in Figure 5 also explains why Horton and Lewis (1992) could calculate binding free energies of rigid protein−protein complexes as sums of two solvation terms with empirical weights and a constant accounting for the translational/rotational entropy loss. Indeed, due to the linear dependence of E^{el}_{r-l} on ΔE^{el}_h, apart from the constants all free energy terms of eq 9 can be expressed as linear combinations of solvent-accessible areas of different atom types. The use of the electrostatic and hydrophobic terms, however, eliminates the need for estimating the coefficients of the two solvation terms from experimental values of binding free energies (Horton & Lewis, 1992).

The results of calculations on the biotin−streptavidin and class I MHC−peptide systems also indicate the importance of taking into account side-chain composition (types of atoms) and not just size in estimating the hydrophobic contribution to the binding free energy. For example, the class I MHC HLA-A2 haplotype has binding pockets that are highly specific for hydrophobic side chains. Because these pockets also have some partially charged groups, charged side chains have a much more favorable electrostatic energy than hydrophobic side chains. When only surface area is used to estimate solvation free energy, the relative free energy advantage of a hydrophobic side chain binding to a hydrophobic pocket is lost and charged side chains are invariably predicted to bind more favorably than hydrophobic side chains (data not shown). On the other hand, the solvation term calculated by the ASP model provides sufficient compensation to predict the preference of the HLA-A2 binding pockets for hydrophobic residues.

Although using the solvation term $\gamma\Delta A_t$ is not generally adequate, the results of Novotny et al. (1989) and Krystek et al. (1993) indicate that in most known antibody−antigen and serine endopeptidase−inhibitor systems this model of solvation provides accurate free energy estimates. Entropic considerations alone are adequate in these systems because electrostatics plays a relatively minor role in the binding region [see Table 12 of Krystek et al. (1993)]. A notable exception is the tissue−kallikrein complex (2KAI; see Table 1), which has side chains with complementary charges at the interface and will be further discussed in the next section.

Consistency and Sensitivity to Parameters. The most critical issue in empirical free energy evaluation is the consistency of various contributions (i.e., energetic, entropic, and solvation) to the free energy. While in principle each term is separately consistent with some thermodynamic scale, they are not necessarily consistent with each other under the common conditions used in free energy evaluation. For example, Wilson et al. (1991) calculated binding free energy by the equation $\Delta G = w_{nb}E_{nb} + w_{solv}\Delta G_{solv}$, where E_{nb} represents the nonbonded (i.e., van der Waals and electrostatic) interaction energy and ΔG_{solv} is the solvation free energy calculated by the ASP model (Eisenberg & McLachlan, 1986). The weighting coefficients w_{nb} and w_{solv} were selected by fitting this model to 40 α-lytic protease−substrate complexes; thus the procedure is less general than the ones presented here or by Novotny et al. (1989). The very different weights ($w_{nb} = 0.031$ and $w_{solv} = 1.98$) show that the two free energy terms are not determined from consistent parameter sets. A possible reason is that the structures for most of these complexes were predicted by homologous extension involving a relatively coarse-grained search for the conformations of some mutated side chains (Wilson et al., 1989). Such calculations tend to yield structures with steric overlaps and inflated values of the van der Waals interaction energy.

Both the method of Novotny et al. (1989) and the one presented here avoid the use of adjustable parameters. Some parameters such as $\epsilon = 4r$ are somewhat arbitrary, but the values adopted are generally accepted and are not optimized for a particular class of proteins. The most arbitrary parameters are likely to be the translational/rotational and cratic entropy terms, but these are assumed to be conforma-

tion-independent and hence affect only the reference state of the free energy.

There is also uncertainty in the values of atomic solvation parameters. We estimated these parameters from the octanol/water partition data of Fauchere and Pliska (1983), which have been adjusted by Sharp et al. (1991) using the Flory–Huggins theory (Flory, 1941, 1942; Huggins, 1941) to account for the differing size of solute and solvent. The Flory–Huggins theory assumes a randomly coiling high polymer under θ solvent conditions (Flory, 1941; Young & Dill, 1990). It therefore cannot provide a completely valid correction term. Alternative treatments in the literature regard the solute as rigid spheres (Meroni et al., 1987) or rigid rods such as linear chains fixed in a fully extended conformation (Ben-Naim & Mazo, 1993). These are also not entirely applicable since the *N*-acetyl amino acid amides considered by Fauchere and Pliska (1983) are flexible. A correct treatment, which is still needed, will undoubtedly be between these extremes and has to address a number of further problems such as the effects of strong, directed interactions between water and polar/charged protein groups (Holtzer, 1992). Nevertheless, De Young and Dill (1990) showed that the use of Flory–Huggins correction was necessary to obtain partition coefficients for benzenes between *n*-alkane hydrocarbons and water that are independent of alkane chain length as anticipated. We also note that without the correction the hydrophobic free energies in Table 1 are increased by a substantial but almost invariant value (14.6 kcal/mol). This affects the constant term in the free energy, which in any case is not well characterized, but leaves the relative binding free energies largely unchanged.

In the case of a rigid ligand, the largest errors in the binding free energy are associated with electrostatic term E^{el}_{r-l}. As discussed earlier, this term can be too large in complexes with many charged side chains in the contact area. In particular, the electrostatic interaction energy is invariably overestimated for 2KAI, although our value is not as unreasonable as the one obtained by Krystek et al. (1993). The difference between the two results is almost 27 kcal (Table 2) and thus is worth comment. This difference is not caused by the use of different solvation models but is related to our use of a later version of the CHARMm potential in refining the crystal structure (CHARMm 22 by Molecular Simulations, Inc., as opposed to CHARMm 19). Because the earlier version allowed somewhat greater flexibility in bond lengths and angles, the minimization shifted the side chain of Lys-15 of the inhibitor (BPTI) close to several negative charges on the kallikrein, particularly to the side chain of Asp-189, resulting in strong electrostatic interactions.

Generalizing the above comments, there are at least two different problems related to electrostatics. One is that the Coulomb potential with fully protonated and deprotonated charge groups becomes unrealistic when charges are too close, and the other is the potential sensitivity of the results to structural deviations. Even with electrostatics treated properly, errors in refinement can introduce large changes in the final results when charge groups are localized. Although we reduced the errors by using a more "rigid" version of the CHARMm potential, the binding free energy calculated for 2KAI is still too negative. The problem may in part be the result of treating acidic and basic groups as fully ionized, whereas their pK values will shift toward neutral due to the Coulombic interaction as they approach one other (Gilson, 1993). In principle, more rigorous treatments of electrostatics such as solving the Poisson–Boltzmann equation (Gilson & Honig, 1987) may improve the binding free energy estimates. More effective treatment of charge–charge interactions will also be crucial to more accurate calculations of the average internal and hydrophobic free energies of the free peptide.

Binding of Peptide Antigens to MHC Class I Receptors. The presentation of antigenic peptides bound to MHC class I molecules is prerequisite to the stimulation of cytotoxic T cells. As such they are prime candidates for vaccine development, both therapeutically and prophylactically.

Table 5: Residue Contributions to Binding Free Energy in Complexes of HLA-A2 with Antigenic Peptides Listed in Table 3

peptide	residue	E^{el}_{bb}	E^{el}_{sc}	E^{el}_{r-l}	ΔG^{res}_{h}	$-T\Delta S_{sc}$	ΔG^{res}
HIV-1 RT	NTER	−7.04	N/A	−7.04	1.67	N/A	−5.37
	I1	−1.23	0.00	−1.23	−4.16	0.89	−4.50
	L2	−3.35	0.00	−3.35	−4.70	0.78	−7.27
	K3	0.98	−0.61	0.37	−1.22	1.94	1.09
	E4	−0.10	−9.98	−10.08	−0.46	1.44	−9.10
	P5	−0.40	0.21	−0.19	−0.42	0.00	−0.61
	V6	0.20	0.00	0.20	−2.77	0.51	−2.06
	H7	−1.37	−0.40	−1.77	−1.47	0.96	−2.28
	G8	−2.19	0.00	−2.19	−0.78	0.00	−2.97
	V9	−0.31	0.00	−0.31	−4.77	0.51	−4.57
	CTER	−11.94	N/A	−11.94	0.43	N/A	−11.51
HTLV-1 Tax	NTER	−9.03	N/A	−9.03	2.01	N/A	−7.02
	L1	−1.11	0.00	−1.11	−4.02	0.78	−4.35
	L2	−3.93	0.00	−3.93	−4.73	0.78	7.88
	F3	0.61	0.00	0.61	−3.84	0.58	−2.65
	G4	−0.48	0.00	−0.48	−0.65	0.00	−1.13
	Y5	−0.63	0.14	−0.49	−0.99	0.37	−1.11
	P6	0.17	0.30	0.47	−0.71	0.00	−0.24
	V7	−0.34	0.00	−0.34	−2.78	0.51	−2.61
	Y8	−2.65	0.06	−2.59	−2.18	0.83	−3.94
	V9	−0.05	0.00	−0.05	−4.71	0.51	−4.25
	CTER	−14.78	N/A	−14.78	0.48	N/A	−14.30
Flu A matrix	NTER	−8.28	N/A	−8.28	2.24	N/A	−6.04
	G1	−1.51	0.00	−1.51	−1.61	0.00	−3.12
	I2	−2.47	0.00	−2.47	−4.69	0.89	−6.27
	L3	0.33	0.00	0.33	−2.96	0.78	−1.85
	G4	−0.38	0.00	−0.38	−0.51	0.00	−0.89
	F5	−0.51	0.00	−0.51	−2.28	0.58	−2.21
	V6	0.28	0.00	0.28	−2.61	0.51	−1.82
	F7	0.38	0.00	0.38	−3.50	0.58	−2.54
	T8	−2.78	−1.81	−4.59	−1.96	1.63	−4.92
	L9	−0.89	0.00	−0.89	−5.65	0.78	−5.76
	CTER	−6.78	N/A	−6.78	0.48	N/A	−6.30
HIV-1 gp120	NTER	−10.59	N/A	−10.59	1.83	N/A	−8.76
	T1	−1.69	0.47	−1.22	−2.51	1.63	−2.10
	L2	−4.40	0.00	−4.40	−4.54	0.78	−8.16
	T3	−0.20	−0.27	−0.47	−1.80	1.63	−0.64
	S4	0.08	−0.31	0.39	−0.44	1.60	1.55
	C5	−0.21	−0.08	−0.29	−0.56	0.55	−0.30
	N6	0.29	0.13	0.42	−0.81	0.39	0.00
	T7	−1.12	−0.10	−1.22	−1.51	1.63	−1.10
	S8	−2.66	1.21	−1.45	−0.92	1.71	−0.66
	V9	−0.19	0.00	−0.19	−4.63	0.51	−4.31
	CTER	−13.95	N/A	−13.95	0.50	N/A	−13.45
Hep B capsid	NTER	−9.49	N/A	−9.49	2.00	N/A	−7.49
	F1	−1.73	0.00	−1.73	−4.40	0.58	−5.55
	L2	−4.31	0.00	−4.31	−4.28	0.78	−7.81
	P3	0.75	−0.19	0.56	−1.94	0.00	−1.38
	S4	−0.10	0.01	−0.09	−0.60	0.37	−0.32
	D5	−0.17	−1.58	−1.75	−1.14	1.25	−1.64
	F6	0.22	0.00	0.22	−2.90	0.58	−2.10
	F7	−0.06	0.00	−0.06	−4.01	0.58	−3.49
	P8	−0.82	0.17	−0.65	−1.97	0.00	−2.62
	S9	−1.86	1.41	−0.45	−0.45	1.71	0.81
	V10	0.03	0.00	0.03	−4.66	0.51	−4.12
	CTER	−14.73	N/A	−14.73	0.51	N/A	−14.22

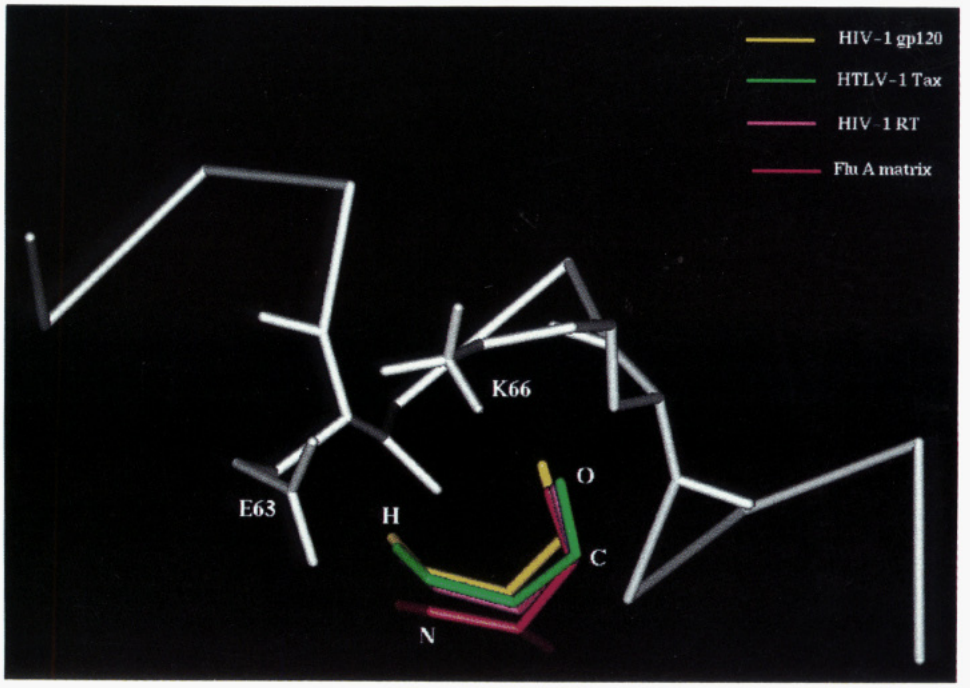

FIGURE 6: Backbones of residue 2 in the X-ray structures of antigenic peptides HIV-1 gp120, HTLV-1 Tax, HIV-1 RT, and Flu A matrix bound to HLA-A2. The nearby backbone portion of the HLA-A2 α_1-helix, as well as its Glu-63 and Lys-66 side chains, is shown in white. Peptides HIV-1 gp120, HTLV-1 Tax, and HIV-1 RT have Leu at position 2. In the Flu A matrix peptide, residue 2 is Ile.

Evidently, candidate peptide vaccines for particular HLA types would be greatly facilitated by an understanding of the sequence correlates of peptide stability. This recognition has led to a number of remarkable discoveries related to sequence motifs. A comparison of the sequences of peptides eluted off any particular class I haplotype invariably finds peptides in the 8−10-residue size range with one or two highly conserved or "anchor" positions (Rotschke et al., 1990; Falk et al., 1991; Jardetzky et al., 1991; Rammensee et al., 1993). With enough such motifs in hand, it would be easy to select from any protein sequence, all peptides having a motif for, and hence being antigenic with, any specified MHC receptor. The experimental motifs, though useful, are neither necessary nor sufficient for binding.

Among the potential applications of the methods developed here is the rank ordering of peptides for any receptor of known structure. The results in this paper, although promising, nevertheless clearly pinpoint computational obstacles, such as the need for treating electrostatics more accurately, that still need to be overcome. In addition to these, the sensitivity of predictions to the accuracy of receptor structure (which would only be known by homology in most cases) has not yet been assessed. Nevertheless, the results represent a step forward and provide some insight into the limitations not only of the computational approaches but also of the experimental approaches as well.

The defining problem in calculating the free energy change for short peptides is an adequate accounting of the contribution or flexibility. Our results (Tables 3 and 4) clearly indicate the importance of that contribution which, in its observed range of 5−10 kcal/mol, represents some 30−50% of the overall free energy change. But its importance lies not just in its magnitude but in its nature, for its contribution is collective; i.e., it cannot be partitioned or factored among the various residues. For the same reason, the change in solvation free energy cannot be cleanly factored; i.e., the terms contributing to its magnitude arise from conformations of the peptide as a whole.

Superimposed upon these collective effects are contributions that are both specific and factorable, i.e., contributions that can be assigned to individual peptide residues. The electrostatic interaction energy E^{el}_{r-l} is fully factorable without omitting any contributions. Between 65% and 75% of E^{el}_{r-l} is contributed by charged end groups, and the rest is almost completely associated with the peptide backbone. Charged side chains also occasionally contribute (see HIV-1 RT). The ΔG^{res}_h values in Table 5 represent the contributions of peptide residues to the solvation free energy. Contributions of MHC residues are not shown. The additional contribution of the MHC to the solvation free energy is not factored, and data are not shown. Similarly, the $T\Delta S_{sc}$ terms in Table 5 represent only the entropy loss of peptide side chains and not those of the receptor.

Both the solvation term ΔG_h and the entropic term $T\Delta S_{sc}$ are sums of contributions from the ligand and the receptor. The receptor contributions are determined by the change in the solvent-accessible surface areas of receptor atoms upon peptide binding. Since all problems described in this section involve the same HLA-A2 receptor, and mutating peptide residues leaves the contact surface of the receptor essentially invariant, only peptide contributions will be factored to individual residues.

Specificity of course resides in particular peptide side chains at particular locations. In HLA-A2, the peptide side chains contribute to binding almost exclusively by hydrophobic interactions. Indeed, the best binders (HTLV-1 Tax, Hep B capsid, and Flu A matrix peptides) have strongly hydrophobic groups of residues around the anchor positions 2 and 9. As we indicated in previous sections, the correct prediction of specificity requires taking account of enthalpic contributions to solvation.

These observations bear on the limitations of any method that would assess stability as a product of uncorrelated or independent contributions from several residues, whether by the method of simple motifs or by more substantial methods. In fact, Parker et al. (1994), who developed a purely empirical method of predicting the stability of peptide binding to HLA-A2, have already noted experimental observations of its limitations. For example, the HIV-1 RT peptide in Table 3 binds to HLA-A2 with a half-life of 3 h, but the value predicted by assuming that each side chain contributes independently to the overall binding stability is 4.8 min (Parker et al., 1994). The predicted short half-life is traceable to the observation that any charged residue at position 3 for nonamers binding HLA-A2 is generally unstable. Therefore, Parker et al. (1994) assumed the existence of overriding favorable interactions with other peptide residues.

In fact, our calculations (Table 5) indicate the presence of a strong favorable electrostatic interaction between the receptor and the side chain of Glu-4 of the peptide. Although the latter is directed sideways and is somewhat out of the groove, it interacts with the nearby side chains of Arg-65 and Lys-66 of the receptor. This interaction is in fact so strong that it would lead to even longer dissociation times were it not for the large change in the peptide energy upon binding (Table 3). The latter is due to interactions between the Lys-3 and Glu-4 side chains in the free ligand, and this again mitigates against an assumption that side chains contribute independently to stability. The interplay of these factors combines to give a reasonable prediction for this peptide (Figure 4).

The results in Table 5 generally corroborate and complement the data from endogenous peptide sequence analyses (Falk et al., 1991; Hunt et al., 1992). The anchor positions in HLA-A2 are 2 and 9. In particular, a mutation of Leu to Ile at position 2 decreases the stability of the peptide 10-fold (Ruppert et al., 1993; Parker et al., 1994). On the basis of the results for the four structures with Leu at position 2 in Table 5, the average contribution to the free energy by Leu at this position (i.e., the sum of electrostatic, hydrophobic, and entropic terms) is −7.78 kcal/mol, whereas Ile at position 2 of the Flu A matrix peptide contributes only −6.27 kcal/mol. This difference, whose origin is the electrostatic interaction energy between the receptor and the peptide backbone, introduces a 12-fold increase in the dissociation constant, in good agreement with observations (Parker et al., 1994).

Both the observation and the result of the calculation are somewhat surprising, since a 10-fold change in stability is ordinarily not expected from so conservative a substitution. The reason can be understood by considering Figure 6, which shows the X-ray structures of the four crystallized nonameric peptides, three with Leu at position 2 and one with Ile. The backbone of the Flu A matrix protein which has an Ile at position 2 is shifted from the other three, which cluster together. This shift places the amide hydrogen of Ile-2 at 3.16 Å from the $O^{\epsilon 2}$ atom of Glu-63, whereas for the other three peptides, the corresponding distances are less than 2.3 Å.

Finally, the anchor residue at position 9 is a Leu in the Flu A matrix peptide, and Leu in this position is known to yield a weaker binding than the C-terminal Val contained in the other four peptides (Parker et al., 1994). The data in Table 5 suggest that this difference is also due to electrostatics, because the Val side chain at position 9 allows the carboxyl end group to occupy a much more favorable position than Leu-9 does.

ACKNOWLEDGMENT

We thank J. Novotny for helpful discussions and for providing us with the programs implementing the binding free energy calculations described in Novotny et al. (1989).

REFERENCES

Adamson, A. W. (1976) *Physical Chemistry of Surfaces*, Wiley, New York.
Avbelj, F. (1992) *Biochemistry 31*, 6290−6297.
Ben-Naim, A., & Mazo, R. M. (1993) *J. Phys. Chem. 97*, 10829−10834.
Bernstein, F., Koetzle, T., Williams, G., Meyer, E., Brice, M., Rodgers, J., Kennard, O., Shimanouchi, T., & Tasumi, M. (1977) *J. Mol. Biol. 112*, 535−542.
Bjorkman, P. J., Saper, M. A., Samraoui, B., Bennet, W. S., Strominger, J. L., & Wiley, D. C. (1987) *Nature 329*, 506−512.
Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., & Karplus, M. (1983) *J. Comput. Chem. 4*, 187−217.
Caflisch, A., Niederer, P., & Anliker, M. (1992) *Proteins 13*, 223−230.
Cherfils, J., & Janin, J. (1993) *Curr. Opin. Struct. Biol. 3*, 265−269.
Chothia, C. (1974) *Nature 248*, 338−339.
Connolly, M. L. (1981) *Molecular Surface Program. Quantum Chemistry Program Exchange*, Indiana University, Bloomington, IN 47401.
De Young, L. R., & Dill, K. A. (1990) *J. Phys. Chem. 94*, 801−809.
Eisenberg, D., & Mclachlan, A. D. (1986) *Nature 319*, 199−203.
Erickson, H. P. (1989) *J. Mol. Biol. 206*, 465−474.
Falk, K., Rotzschke, O., Stevanovic, S., Jung, G., & Rammensee, H.-G. (1991) *Nature 351*, 290−296.
Fauchere, J.-L., & Pliska, V. (1983) *Eur. J. Med. Chem. 18*, 369−375.
Flory, P. J. (1941) *J. Chem. Phys. 9*, 660−661.
Flory, P. J. (1942) *J. Chem. Phys. 10*, 51−61.
Gilson, M. K. (1993) *Proteins 15*, 266−282.
Gilson, M. K., & Honig, B. H. (1988) *Proteins 3*, 32−52.
Goodsell, D. S., & Olson, A. J. (1990) *Proteins 8*, 195−202.
Green, N. M. (1966) *Biochem. J. 101*, 774−780.
Green, N. M. (1975) *Adv. Protein Chem. 29*, 85−143.
Green, N. M. (1990) *Methods Enzymol. 184*, 51−67.
Hart, T. N., & Read, N. J. (1992) *Proteins 13*, 206−222.
Hendrickson, W. A., Pahler, A., Smith, J. L., Satow, Y., Merritt, E. A., & Phizackerley, R. P. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 2190−2194.
Hermann, R. B. (1972) *J. Phys. Chem. 76*, 2754−2759.
Holtzer, A. (1992) *Biopolymers 32*, 711−715.
Horton, N., & Lewis, M. (1992) *Protein Sci. 1*, 169−181.
Huber, R., & Bode, W. (1978) *Acc. Chem. Res. 11*, 114−122.
Huggins, M. L. (1941) *J. Chem. Phys. 9*, 440−440.
Jardetzky, T. S., Lane, W. S., Robinson, R. A., Madden, D. R., & Wiley, D. C. (1991) *Nature 353*, 326−329.
Kauzmann, W. (1959) *Adv. Protein Chem. 14*, 1−64.
Krystek, S., Stouch, T., & Novotny, J. (1993) *J. Mol. Biol. 234*, 661−679.
Leach, A. R., & Kuntz, I. D. (1992) *J. Comput. Chem. 13*, 730−748.
Lee, B., & Richards, F. M. (1971) *J. Mol. Biol. 55*, 379−400.

Madden, D. R., Garboczi, D. N., & Wiley, D. C. (1993) *Cell 75*, 693−708.
Meroni, A., Pimpinelli, L., & Reatto, L. (1987) *J. Chem. Phys. 87*, 3644−3646.
Mezei, M., & Beveridge, D. L. (1986) Computer Simulations and Biomolecular Systems, *Ann. N.Y. Acad. Sci. 494*, 1−23.
Miyamoto, S., & Kollman, P. A. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 8402−8406.
Miyamoto, S., & Kollman, P. A. (1993) *Proteins 16*, 226−245.
Murphy, K. P., Xie, D., Garcia, K. C., Amzel, L. M., & Freire, E. (1993) *Proteins 15*, 113−120.
Nicholls, A., Sharp, K. A., & Honig, B. (1991) *Proteins 11*, 281−296.
Novotny, J., Bruccoleri, R. E., & Saul, F. A. (1989) *Biochemistry 28*, 4735−4749.
Parker, K. C., Bednarek, M. A., Hull, L. K., Utz, U., Cunningham, B., Zweerink, H. J., Biddison, W. E., & Coligan, J. E. (1992a) *J. Immunol. 149*, 2580−3587.
Parker, K. C., Dibrino, M., Hull, L., & Coligan, J. E. (1992b) *J. Immunol. 149*, 1896−1904.
Parker, K. C., Bednarek, M. A., & Coligan, J. E. (1994) *J. Immunol. 152*, 163−175.
Pickett, S. D., & Sternberg, M. J. E. (1993) *J. Mol. Biol. 231*, 825−839.
QUANTA Software, Release 3.0 (1990) Polygen Corp., Waltham, MA.
Reynolds, J. A., Gilbert, D. B., & Tanford, C. (1974) *Proc. Natl. Acad. Sci. U.S.A. 71*, 2925−2927.
Reynolds, C. A., King, P. M., & Richard, W. G. (1992) *Mol. Phys. 76*, 251−275.
Rosenfeld, R., Zheng, Q., Vajda, S., & DeLisi, C. (1993) *J. Mol. Biol. 234*, 515−521.
Rotzschke, O., Falk, K., Deres, K., Schild, K., Norda, M., Metzger, J., Jung, G., & Rammensee, H. G. (1990) *Nature 348*, 252−254.
Ruppert, J., Sideney, S., Celis, E., Kubo, R. T., Grey, H. M., & Sette, A. (1993) *Cell 74*, 929−937.
Sezerman, U., Cornette, J., Vajda, S., & Delisi, C. (1993) *Protein Sci. 2*, 1827−1846.
Sharp, K. A., Nicholls, A., Fine, R. F., & Honig, B. (1991a) *Science 252*, 106−109.
Sharp, K. A., Nicholls, A., Friedman, R., & Honig, B. (1991b) *Biochemistry 30*, 9686−9697.
Shoichet, B. K., & Kuntz, I. D. (1991) *J. Mol. Biol. 221*, 327−346.
Shrake, A., & Rupley, J. A. (1973) *J. Mol. Biol. 79*, 351−371.
Smith, K. C., & Honig, B. (1994) *Proteins 18*, 119−132.
Stoddar, B. L., & Koshland, D. E., Jr. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 1146−1153.
Vajda, S. (1993) *J. Mol. Biol. 229*, 125−145.
van Gunsteren, W. F., & Berendsen, H. J. C. (1990) *Angew. Chem., Int. Ed. Engl. 29*, 992−1023.
Weber, P. C., Ohlendorf, J. J., & Salemme, F. R. (1989) *Science 243*, 85−88.
Weber, P. C., Wendoloski, J. J., Pantoliano, M. W., & Salemme, F. R. (1992) *J. Am. Chem. Soc. 114*, 3197−3200.
Wilson, C., Mace, J. E., & Agard, D. A. (1991) *J. Mol. Biol. 220*, 495−506.
Wilson, C., Gregoret, L. M., & Agard, D. A. (1993) *J. Mol. Biol. 229*, 996−1006.