# Optimization of the GB/SA Solvation Model for Predicting the Structure of Surface Loops in Proteins

**Agnieszka Szarecka and Hagai Meirovitch***

*Department of Computational Biology, University of Pittsburgh School of Medicine, Suite 3064, BST 3, 3501 Fifth Avenue, Pittsburgh, Pennsylvania 15260*

*Received: October 10, 2005; In Final Form: December 1, 2005*

Implicit solvation models are commonly optimized with respect to experimental data or Poisson−Boltzmann (PB) results obtained for small molecules, where the force field is sometimes not considered. In previous studies, we have developed an optimization procedure for cyclic peptides and surface loops in proteins based on the *entire* system studied and the specific force field used. Thus, the loop has been modeled by the *simplified* solvation function $E_{tot} = E_{FF}$ ($\epsilon = 2r$) + $\Sigma_i \sigma_i A_i$, where $E_{FF}$ ($\epsilon = nr$) is the AMBER force field energy with a distance-dependent dielectric function, $\epsilon = nr$, $A_i$ is the solvent accessible surface area of atom $i$, and $\sigma_i$ is its atomic solvation parameter. During the optimization process, the loop is free to move while the protein template is held fixed in its X-ray structure. To improve on the results of this model, in the present work we apply our optimization procedure to the physically more rigorous solvation model, the generalized Born with surface area (GB/SA) (together with the all-atom AMBER force field) as suggested by Still and co-workers (*J. Phys. Chem. A* **1997**, *101*, 3005). The six parameters of the GB/SA model, namely, $P_1-P_5$ and the surface area parameter, $\sigma$ (programmed in the TINKER package) are reoptimized for a "training" group of nine loops, and a best-fit set is defined from the individual sets of optimized parameters. The best-fit set and Still's original set of parameters (where Lys, Arg, His, Glu, and Asp are charged or neutralized) were applied to the training group as well as to a "test" group of seven loops, and the energy gaps and the corresponding RMSD values were calculated. These GB/SA results based on the three sets of parameters have been found to be comparable; surprisingly, however, they are somewhat inferior (e.g, of larger energy gaps) to those obtained previously from the simplified model described above. We discuss recent results for loops obtained by other solvation models and potential directions for future studies.

## Introduction

**Interest in Surface Loops and the Difficulty in Predicting their Structure.** A surface loop in a protein is a chain segment connecting two secondary structure elements, which generally protrudes into the solvent and thus is expected to be relatively flexible, as indeed has been found by multidimensional nuclear magnetic resonance (NMR) experiments. In many cases, this flexibility is also reflected in X-ray crystallography data in terms of large B factors[1] or complete disorder. Surface loops take part in protein−protein and protein−ligand interactions, where their flexibility in many cases is essential for these recognition processes. For example, the conformational change between a free and a bound antibody demonstrates the flexibility of the antibody combining site, which typically includes hypervariable loops; this provides an example of *induced fit* as a mechanism for antibody−antigen recognition (e.g., see refs 2 and 3). Alternatively, the *selected-fit* mechanism has been suggested, where the free loop interconverts among different states, and one of them is selected upon binding.[4] Dynamic NMR experiments[5] and molecular dynamics (MD) simulations[6] of HIV protease have found a strong correlation between the flexibility of certain segments of the protein and the movement of the flaps (that cover the active site) upon ligation.[7] Loops are known to form "lids" over active sites of proteins, and mutagenesis experiments show that residues within these loops are crucial for substrate binding or enzymatic catalysis; again, these loops are typically flexible (see the review by Fetrow in ref 8).

Prediction of loop structures by computational methods is important in homology modeling, where a framework of unconnected homologous segments is initially created and the structure of the loops connecting these segments has to be subsequently determined. For long loops this is an unsolved problem to date.[9−12] Prediction of loop structures also constitutes a challenge in protein engineering, where a loop undergoes mutations, insertions, or deletions of amino acids. Studying the flexibility of loops by experimental methods is not straightforward, and theoretical analysis by molecular modeling techniques is expected to clarify the picture.

The interest in surface loops has yielded extensive theoretical work where one avenue of research has been the classification of loop structures.[13−21] However, to understand various recognition mechanisms such as those mentioned above, it is mandatory to be able to predict the structure (or structures) of a loop by theoretical/computational procedures, which is not a trivial task because of the irregular structures of loops, their flexibility, and exposure to the solvent. Loop structures are commonly predicted by either a comparative modeling approach based on known loop conformations from the protein data bank (PDB),[22,23] or an energetic approach; also, methods exist that are hybrids of these two approaches. Because of the lack of sufficiently large databases, only short loops (up to five residues) could be treated effectively by comparative modeling,[24−29] while hybrid methods are effective up to nine residues.[24,26,30−33] With the energetic

* Corresponding author. E-mail: hagaim@pitt.edu. Tel: 412-648-3338. Fax: 412-648-3163.

10.1021/jp055771+ CCC: $33.50 
Published on Web 01/21/2006

approach, loop structures are generated by conformational search methods (simulated annealing, bond relaxation algorithm, and others) subject to the spatial restrictions imposed by the *known* 3D structure of the rest of the protein (the template). The quality of the prediction depends on the quality of the loop−loop and loop−template interaction energy, the modeling of the solvent, and the extent of conformational search applied.[34−44] An extensive discussion, references, and background material on loops appear in our previous work, denoted here as papers I (ref 45) and II (ref 46).

In the energetic approach, modeling of the solvent is of special importance. In some of the earlier studies the solvation problem was not addressed at all, whereas others used only a distance-dependent dielectric function ($\epsilon = r$). Better treatments of solvation were applied by Moult and James[35] and Mas et al.[47] A systematic comparison of solvation models was first carried out by Smith and Honig,[48] who tested the $\epsilon = r$ model against results obtained by the finite difference Poisson−Boltzmann (FDPB) calculation including a hydrophobic term; the implicit solvation model of Wesson and Eisenberg[49] with $\epsilon = r$ was also studied by them. Later, the generalized Born surface area (GB/SA) model[50] was applied to loops of ribonuclease (RNase) A[51] and has been found by Blundell's group to discriminate better than other models between the native loop structures and close to native "decoy" structures.[52,53] Very recently, an extensive study of loops was carried out by Jacobson et al.[54] who used the surface GB[55] and a nonpolar solvation model[56] (SGB-NP) with the OPLS force field.[57] Zhang et al.[58] have tested their knowledge-based statistical potential, DFIRE (distance-scaled, finite ideal gas reference state) by applying it to the loop sets studied in refs 52−53 and 54 (see the Results and Discussion section). Another interesting loop prediction algorithm has been suggested by Xiang et al.[59] Finally, we mention our loop studies in papers I and II, which will be discussed in detail later. However, more work is needed to compare the quality of the various models for loops and other systems.

**Statistical Mechanics Methodology for Treating Flexibility.** The foregoing discussion indicates that, to date, the energetic approach is the best way for predicting the structure of *large* loops in homology modeling and protein engineering. It also constitutes the only alternative for studying *intermediate flexibility*, where a loop populates several microstates in equilibrium (see below). Recently, we have developed a statistical mechanics methodology for treating intermediate flexibility (most suitable for implicit solvation models), which was applied initially to peptides,[60−65] and in papers I and II, also to surface loops.[66−68] The first step is to carry out an extensive conformational search using our local torsional deformation (LTD) method,[45,46 60,69,70] from which the global energy minimum (GEM) loop structure and low-energy minimized structures within 2−3 kcal/mol above GEM are identified; a subgroup of them that are *significantly different* are then selected where each becomes a "seed" for a local Monte Carlo (MC) or MD simulation that spans its vicinity (this local region is called the *microstate*). Finally, the free energies of the most stable microstates are obtained (with the local states method[71,72] or the hypothetical scanning MC method[73,74]), which lead to the populations and to weighted averages of physical quantities that are compared with the experiment.[61,64,65] Developing a reliable solvation energy function is mandatory and thus is the aim of this paper (as has been the aim of papers I and II).

**Previous Optimization of a Simplified Solvation Model.** Because explicit solvent, the most accurate model, is computationally expensive, we have chosen to study *initially* a relatively simple implicit solvation model defined by eq 1, which was applied to cyclic peptides in DMSO, and in papers I and II to loops in water

$$E_{\mathrm{tot}} = E_{\mathrm{FF}}(\epsilon = nr) + E_{\mathrm{solv}} = E_{\mathrm{FF}}(\epsilon = nr) + \sum_i \sigma_i A_i \quad (1)$$

$E_{\mathrm{FF}}$ is the force field energy, $A_i$ is the structure dependent solvent accessible surface area of atom $i$, and $\sigma_i$ is the atomic solvation parameter (ASP); $\epsilon = nr$ is a distance-dependent dielectric function, where $n$ is a parameter. Even with such a simplified model, treatment of loops is feasible only for a relatively small template that typically consists of those atoms that are located within 10 Å from any loop atom in a specific loop structure; the template atoms are fixed in their known X-ray structure, whereas the loop is free to move. $E_{\mathrm{tot}}$ includes the loop−loop and loop−template energy, while the template−template interactions are ignored. With this model, the conformational search, the identification of the most stable microstates, and the calculation of their free energy is considerably easier than with explicit solvent. Therefore, most of the loop studies in the literature are based on implicit solvation models with relatively small number of exceptions where explicit models were used (e.g., refs 27, 75, and 76).

Equation 1 is not new and has been used in many previous studies, where the ASPs for a protein have been commonly determined from the free energy of transfer of small molecules from the gas phase to water.[49,77] However, it is not clear to what extent ASPs derived for small molecules are suited for the protein environment. Also, these sets of ASPs were used with various force fields, in most cases without further calibration (see discussions in refs 60 and 63, and in references therein). Recent studies based on various solvation potentials, $E_{\mathrm{solv}}$, including our results in papers I and II, support these reservations.[48,51] This problem was first recognized by Schiffer et al.[78] and then by Fraternali and van Gunsteren.[79] Optimization of solvation models with respect to a force field has now become a common practice.

We have developed a procedure for optimizing the parameters of implicit solvation models that to a large extent is free of the limitations discussed above. This procedure was applied first to cyclic peptides and recently to loops modeled by eq 1; in an attempt to further improve the latter results, our main objective in this paper is to apply this procedure to the GB/SA model of Still and co-workers,[50] which relies on stronger theoretical grounds than eq 1. We shall compare results for loops obtained in papers I and II (using eq 1) to the GB/SA results, and will study eq 1 again, where $\epsilon = nr$ is replaced by more complex dielectric functions. Because the general features of the optimization method apply to any model, we discuss them with respect to eq 1.

Thus, for a given loop the optimized ASPs and $n$ are those for which the known X-ray loop structure becomes the GEM structure. This definition, however, turns out to be too strict and in papers I and II we argue that it can be relaxed; thus, an energy difference (the energy gap) of up to 2−3 kcal/mol is allowed between the GEM and the energy of the native optimized structure (NOS) (obtained by local energy minimization of the known X-ray loop structure using the optimized parameters; a more precise definition of NOS will be given later. $E_{\mathrm{FF}}$ (eq 1) is defined by the all-atom AMBER[80] force field that for loops has been found to perform better than other force fields (see paper I). The optimization is based on an extensive conformational search using LTD; this program has been implemented within the molecular mechanics/molecular dynam-

ics program TINKER.[81] For the optimized sets of ASPs (denoted $\sigma_i^*$) and the optimal $n = 2$, the energy gap, $\Delta E_{tot}^m(n,\sigma_i^*)$ is defined by

$$\Delta E_{tot}^m(n,\sigma_i^*) = E_{tot}^{NOS}(n,\sigma_i^*) - E_{tot}^m(n,\sigma_i^*) \quad (2)$$

where $E_{tot}^m(n,\sigma_i^*)$ is the lowest minimized energy obtained, which is assumed to be the GEM. $E_{tot}^{NOS}(n,\sigma_i^*)$ is the minimized energy of NOS based on the optimal parameters. Thus, unlike the conventional parametrization of eq 1 that relies on free energy of transfer data of *small* molecules, our derivation of the ASPs depends on the force field used and is based on the energy of the *entire* loop in the protein environment.

Our aim is to derive ASPs for the solution environment, where the side chains of a surface loop, and to a lesser extent also the backbone, typically exhibit intermediate flexibility.[82,83] It should be noted, however, that our optimization is carried out with respect to a *single* X-ray crystal structure, where some aspects of its flexibility are only expressed by elevated B factors. This problem may be alleviated as high-resolution X-ray structures become available, which enables one to extract information about side chain rotamers and their populations.[84,85] Notice also that the derivation of the ASPs is based on the minimized energies, thus ignoring the flexibility (i.e., entropy) of the microstates. The first step to eliminating this limitation was done in paper II, where differences in the *free energy* for three loops were calculated. $E_{tot}$ is a free energy function that depends on the temperature (through the $\sigma_i$) but will be referred to as energy. It should also be emphasized that the ASPs are derived only for surface loops that protrude into the solvent due to strong hydrophilic interactions. Indeed, the individual sets of ASPs optimized in papers I and II are mostly negative (hydrophilic), even those of carbon (in contrast to the positive ASP obtained by Wesson and Eisenberg[49] (see discussion in paper II).

From initial studies in paper II it became evident that for highly charged loops the Coulombic interactions are too strong, leading to large energy gaps (in some cases of ~20 kcal/mol); therefore, in all calculations the charges of Arg, Lys, His, Asp, and Glu were neutralized. Individual sets of ASPs were optimized for a diverse (training) group of 12 surface loops of 5−12 residues from different proteins. The extent of similarity among the optimized individual sets enabled defining a reasonable best-fit set of ASPs, which was tested on the training group as well as on an additional (test) group of eight loops. The results for eq 1 where found to be much better than those obtained with the force field [$E_{FF}$ ($\epsilon = 2r$) alone. The root-mean-square deviations (RMSD) of the GEM structures from the corresponding NOS were found in most cases better than those obtained by other methods. However, the energy gaps in many cases were above 3 kcal/mol because of strong electrostatic interactions; this has motivated us to study the GB/SA model that treats these interactions in a more rigorous way than eq 1.

## Theory and Methods

In this section we describe the GB/SA model and the LTD method and provide specific details about the methodology and the calculations.

**GB/SA Solvation Model.** Several versions of the GB/SA model are currently available, where their parameters are commonly optimized against properties of small molecules, experimentally determined solvation energies, or free energies obtained by the Poisson−Boltzmann (PB) equation; in general, the more complex models show better agreement with PB at the expense of an increase in computer time.[50,51,55,56,86−96] With the model of Still and co-workers[50,86] (implemented in TINKER) the solvation energy, $E_{sol}$, consists of an electrostatic polarization energy term, $E_{pol}$, and a nonpolar (hydrophobic) energy component, $E_{hyd} = \Sigma A_i\sigma_i$ (compare with eq 1), thus

$$E_{sol} = E_{pol} + \Sigma A_i\sigma_i \quad (3)$$

where $E_{sol}$ is a free energy term, which, as before, in most cases will be referred to as energy. The total electrostatic energy, $E_{es}$, of the system (in kcal/mol) is

$$E_{es} = 332\sum_{i<j}\frac{q_i\,q_j}{\epsilon_{in}r_{ij}} + E_{pol} = 332\sum_{i<j}\frac{q_i\,q_j}{\epsilon_{in}r_{ij}} - 166\left(\frac{1}{\epsilon_{in}} - \frac{1}{\epsilon_w}\right)\sum_{i,j}\frac{q_i\,q_j}{f_{GB}} \quad (4)$$

where

$$f_{GB} = [r_{ij}^2 + \alpha_i\alpha_j\exp(-r_{ij}^2/k\alpha_i\alpha_j)]^{1/2}$$

and $q_i$ is the charge of atom $i$, $r_{ij}$ is the distance (Å) between atoms $i$ and $j$, $\alpha_i$ is the Born radius of atom $i$, and $k$ is a factor that is taken as 4 in ref 50. $E_{pol}$ is the electrostatic component of the free energy of transfer of a molecule with an interior dielectric constant, $\epsilon_{in}$, from vacuum to a continuum medium (water) of dielectric constant $\epsilon_{w.}$ The total energy, $E_{tot}$, is

$$E_{tot} = E_{FF} + E_{pol} + E_{hyd} \quad (5)$$

where $E_{FF}$ is the energy of the all-atom AMBER94 force field,[80] which includes the first term of $E_{es}$ (eq 4); AMBER94 is chosen to be consistent with eq 1 studied in papers I and II. Notice that in TINKER $E_{hyd}$ is defined as a product of a *single* parameter $\sigma$ and the total surface area of the solute calculated with a spherical solvent molecule (water) of radius 1.4 Å.

The heart of the GB/SA model is the calculation of the $\alpha_i$ values, which in the work of Still and co-workers are defined by a function depending on five parameters, $P_1$−$P_5$ (see ref 50); thus

$$\alpha_i = -166/G'_{pol,i} \quad (6)$$

where

$$G'_{pol,i} = \frac{-166}{R_{vdW-i} + \phi + P_1} + \sum^{stretch}\frac{P_2V_j}{r_{ij}^4} + \sum^{bend}\frac{P_3V_j}{r_{ij}^4} + \sum^{nonbonded}\frac{P_4V_j\mathrm{CCF}}{r_{ij}^4} \quad (7)$$

and $\phi = -0.09$ Å is a dielectric offset. $r_{ij}$ = distance between atoms $i$ and $j$(Å), $V_j$ = volume of atom $j$ (Å$^3$), $R_{vdW-i}$ = van der Waals radius of atom $i$ (Å), $P_1$ = single atom scaling factor, $P_2 = 1$, 2 scaling factor, $P_3 = 1$, 3 scaling factor, $P_4$ 1, ≥ 4 = scaling factor, $P_5$ = soft cutoff parameter, and CCF = close contact function for 1 ≥ 4 interactions where

$$\mathrm{CCF} = 1.0 \quad \text{if} \quad \left(\frac{r_{ij}}{R_{vdW-i} + R_{vdW-j}}\right)^2 > \frac{1}{P^5} \quad (8)$$

otherwise

$$\mathrm{CCF} = \left\{0.5\left[1.0 - \cos\left\{\left(\frac{r_{ij}}{R_{\mathrm{vdW}-i} + R_{\mathrm{vdW}-j}}\right)^2 P_5 \pi\right\}\right]\right\}^2$$

We optimize parameters $P_1-P_5$ and $\sigma$.

**LTD Method.** The local torsional deformation (LTD)[60,69] method has been described in detail before. Here we discuss only its main features. This is a conformational search procedure for cyclic molecules and protein loops modeled by a force field with flexible bond lengths and angles. An LTD simulation starts from an arbitrary energy minimized loop structure, *i*, with energy $E_i^0$; *i* is then distorted by a single or several *local* torsional rotations along the chain followed by energy minimization. The resulting conformation, *j* (with minimized energy $E_j^0$), is accepted according to the Metropolis transition probability, $p_{ij}$

$$p_{ij} = \min\,(1, \exp[-(E_j^0 - E_i^0)/k_{\mathrm{B}}T^*]) \qquad (9)$$

where the accepted structure is deformed again and the process continues. This Monte Carlo minimization procedure,[97] is a "selection procedure" that efficiently directs the search toward the low-energy region in conformational space. Notice that $T^*$ is not a usual temperature but a parameter that affects the efficiency of the process.[98] In most of our runs, $T^*$ was changed every 50 Monte Carlo (LTD) steps by 10 K from 200 to 1000 K and vice versa. The coordinates and energies of all of the energy minimized structures (including those that were rejected through eq 9) were stored in a file for further analysis.[60]

The local backbone rotations are described elsewhere.[60,69] Typically, in each LTD step several independent but significant such rotations (determined randomly) are carried out along the chain, and therefore energy barriers are crossed efficiently. These local conformational changes are especially important in a dense protein environment to reduce the chance for creating undesired loop−template entanglements. Notice that together with the backbone angles side-chain dihedrals are selected randomly as well and they are changed at random (*but not locally*). Thus, the whole loop is treated at once, in contrast to procedures used by others and discussed in papers I and II. The present implementation of LTD is exactly the same as that applied to the cyclic hexapeptide described in detail in ref 60. LTD has been found to be significantly more efficient than simulated annealing.[69]

It should be pointed out that while Monte Carlo minimization (thus LTD) is a stochastic procedure, the chance of finding the GEM is higher if the search starts from a conformation that is similar to the GEM structure than from a distant conformation. Therefore, we start all of the LTD runs from the native loop structures (NOS), which are not expected to differ significantly from the corresponding GEMs. This choice would lead to the expected increase in the search efficiency only if the loop does not get trapped in the starting microstate, which was verified by the relatively large RMSD values (up to ~6 Å) obtained for the trajectories of the generated loops (meaning that a significant part of conformational space was sampled) and the fact that in many cases the energy was decreased significantly. Finally, the energy is minimized by the L-BFGS procedure,[99] which (like the LTD program) has been incorporated in TINKER.

**Loops Studied and Modeling Issues.** It should first be pointed out that the backbone structure of a stretched loop will be predicted correctly by all conformational search methods (see discussion in paper II). Therefore, as in papers I and II we obtained for each loop the ratio, $R$ = [length of a completely stretched (extended) loop/distance between its ends], where these lengths are calculated between the $C^\alpha$ atoms of the first and last residues of the loop. The length (in Å) of the extended structure is calculated using the expressions, $6.046(n/2 - 1) + 3.46$ and $6.046(n - 1)/2$ for an even and odd number, $n$ of residues, respectively; the factors 6.046 and 3.46 Å are taken from Flory[100] (Chapter VII, p 251). To a large extent, $R$ reflects the conformational freedom of the loop backbone and partially also of the side chains, the larger is $R$ the higher the flexibility (which is also determined by the surrounding template and sequence of residues).

To be able to compare the performances of GB/SA and eq 1, we have chosen the same training group of loops studied in paper II, besides the two loops of BPTI [(6−12) and (18−24)] and the loops (119−125) of myoglobin that are extremely stretched ($R$ = 1, 1, and 1.1, respectively). We added to this group the loop (64−71) of RNase A (loop 1) and for each of the nine loops of this group an individual set of parameters were optimized. Again, as in paper II, for each of these loops an individual set of parameters were optimized; the extent of similarity among these sets enabled us to define a reasonable best-fit set of ASPs, which was tested on the training group as well as on an additional test group of seven loops that were also studied in paper II; these groups of loops, the related proteins, and template sizes appear in Table 1.

The 3D structures of the proteins of the training group (taken from the PDB) were all determined with 2 Å resolution or less, except for that of antibody McPC603 that was obtained with 2.7 Å resolution. These loops range in size from 5 to 12 amino acid residues, and all of them are predominantly hydrophilic, that is, polar or charged. It should be pointed out that the coordinates of the side chain atoms of the highly charged loops of acidic FGF (two charged residues) and AK (three charged residues) were obtained with elevated B factors, 47−88 for AK, and 50−100 for chain B of acidic FGF (see detailed discussion in paper II). These large B factors suggest that the side chains might populate several rotamers, but no analysis of such populations is available [Müller and Schulz do not determine dihedral angles if the B factors of the involved atoms are 60 and above,[101] whereas others adopt even a smaller value of 40 (J. Rosenberg, private communication)]. Obviously, this uncertainty in the coordinates of the loops will be reflected in the reliability of the corresponding optimized sets of parameters. The optimized parameters might also be affected by the existence of more than one molecule in the unit cell as is the case for AK and acidic FGF, which have two and four molecules in the unit cell, respectively. Indeed, in paper II we have found that for FGF the B factors and energy gaps of loop 90−94 in molecules B and C are different due to different environments. In the present study we have taken into consideration molecule B only. The optimized parameters might also be affected by close molecules in neighbor cells. However, we have not investigated this point.

The number of atoms (including hydrogens) of the training group ranges from 84 (acidic FGF) to 175 (the 12-residue loop of the antibody; see Table 1). The template is defined by the following procedure. First, hydrogen atoms are added to the PDB X-ray structure by the program TINKER. In the second step, to remove possible atomic overlaps, the energy of the protein is minimized using the AMBER potential [$E_{\mathrm{FF}}$ ($\epsilon$ = 1), eq 1] with an additional harmonic restraint of 5 kcal/mol/Å$^2$ applied to each atomic position. This minimized structure is the native optimized structure (NOS), mentioned earlier, which can deviate from the PDB structure by an all-heavy-atom RMSD of no more than ~0.15 Å. Most templates include any nonloop atom with a distance smaller than 10 Å from at least one loop

**TABLE 1: Proteins and the Corresponding Loops and Templates[a]**

| protein | loop | sequence | *R* | no. atoms (loop) | no. atoms (template) | radius (Å) (template) |
|---|---|---|---|---|---|---|
| | | training group | | | | |
| RNase A (1rat) | loop 3, 89−97 (9) | SS**K**YPNCAY | 2.8 | 133 | 726 | 10 |
| RNase A (1rat) | loop 1 64−71 (8) | AC**K**NGQTN | 3.2 | 107 | 745 | 10 |
| acidic fibroblast (FGF) (2afg) | 90−94 (5) | **E**E**N**H**Y** | 2.3 | 84 | 700 | 10 |
| adenylate kinase (AK) (4ake) | 73−80 (8) | AQ**ED**C**R**NG | 2.1 | 112 | 856 | 10 |
| peptidase (5cpa) | 205−213 (9) | PYGYTTQSI | 3.5 | 138 | 1109 | 10 |
| antibody, McPC603 (1mcp) | loop 1, L26−L37 (12) | SQSLLNSGNQ**K**N | 2.5 | 175 | 893 | 9 |
| antibody, McPC603 (1mcp) | loop 2, H102−H109 (8) | YYGSTWYF | 3.7 | 139 | 1492 | 9 |
| penicillopepsin (3app) | 129−137 (9) | INTVQPQSQ | 2.7 | 139 | 999 | 9 |
| proteinase (2apr) | 202−210 (9) | ATVGTSTVA | 4.8 | 112 | 804 | 9 |
| | | test group | | | | |
| ser-proteinase (2ptn) | 143−151 (9) | NT**K**SSGTSY | 4.9 | 117 | 809 | 9 |
| proteinase (2apr) | 188−196 (9) | I**D**NS**R**GWWG | 4.5 | 143 | 1270 | 9 |
| proteinase (2apr) | 128−137 (10) | **D**TITTV**R**GV**K** | 4.3 | 158 | 1145 | 9 |
| peptidase (5cpa) | 244−250 (7) | ITTIYQA | 2.7 | 114 | 1010 | 9 |
| RNase H (2rn2) | 57−63 (7) | **E**AL**K**E**H**C | 1.6 | 110 | 929 | 9 |
| antibody (1mcp) | 56L−62L (7) | GAST**RE**S | 1.3 | 93 | 1007 | 9 |
| antibacterial protein (1noa) | 25−30 (6) | GLQAGT | 1.3 | 74 | 536 | 9 |

[a] *R* is the ratio between the length of the stretched (extended) loop and the distance between the $C^{\alpha}$ of the first and last residues of the loop. The charged residues are bold-faced.

atom (in NOS) together with all the other atoms belonging to the same residue. However, for some of the larger proteins distances smaller than 10 Å were used to keep the template size manageable. The smaller cutoff distance is justified in light of our finding (paper II) that decreasing the distance from 10 to 7 changed the energy only slightly ($\leq$ 1 kcal/mol), suggesting that the effect on energy *differences* between two structures would be small. The template sizes in Table 1 range from 700 (acidic FGF) to 1492 (antibody, loop 2), which are larger than their counterparts in paper II due to larger radii.

The test group (see Table 1) includes seven of the eight loops studied in paper II, where loop 1 of RNase A was transferred to the training set. All of them are unstretched solvent-exposed surface loops with B factors smaller than 40, except for the loop of ser-proteinase, where all the coordinates are given but seven outer atoms of side chains have zero electron density. For all of these loops the templates have been defined with a radius of 9 Å.

**More Details about the Optimization Procedure.** TINKER assigns the hydrogen atoms to the PDB structure by a prescription that does not optimize their positions with respect to the energy; therefore, in paper I it was necessary to optimize the orientations of the OH and NH vectors of NOS and the template. This is carried out by a Monte Carlo minimization procedure, where the polar vectors are rotated by LTD while each nonrotatable atom is restrained to its NOS position by a harmonic potential of 0.15−0.40 kcal/mol/Å$^2$ (see Appendix C of paper I). These optimizations of the polar hydrogen networks [using $E_{FF}$ ($\epsilon = 10$)], carried out in paper II and here, lead to NOS structures that deviate by RMSD ~0.2 Å from the PDB loop structures; these structures, denoted NOS1 (to be distinguished from NOS2 defined later), are considered to be the correct (experimental) ones against which the RMSD values of the structures are calculated.

As for the ASPs, in the GB/SA optimizations the charges of Arg, Lys, His, Asp, and Glu, and the end groups of the protein are neutralized to decrease the effect of the electrostatic interactions (see details in paper II); notice, however, that these interactions are still significant because of large dipole moments. Also, for all of the loops we carry out LTD runs based on Still's original (standard) parameters with neutralized as well as charged Arg, Lys, His, Asp, and Glu.

The optimization of the parameters is based on a multistage search for low-energy minimized structures carried out with LTD, as described in detail in Appendix B of paper I. In short, for each loop the first stage is a conformational search run of ~3000 energy minimizations based on Still's original set of parameters (denoted $P_1-P_5$). From this sample we define a subgroup of 500−800 *significantly different* structures (according to the variance criterion that at least one dihedral angle differs by 60° or more) with minimized energies within a ~7 kcal/mol range above the GEM (assumed here to correspond to the lowest minimized energy structure generated). NOS1 is added to this group as well. At this stage parameter $P_1$ is optimized ($P_2-P_5$ are kept fixed) by changing its value and minimizing the energy of the above group of structures to find the value ($P'_1$) that leads to the smallest energy gap between GEM and the minimized NOS1 (eq 2). $P'_1$ is a temporary optimized value that is kept constant when $P_2$ is optimized in the same way. However, the subgroup of structures might not remain of low energy for the set $P'_1$, $P'_2$, $P_3-P_5$. Therefore, a new LTD run based on the latter values is performed and a new subgroup is determined, which is used in the optimization of $P_3$ and so forth. After optimizing $P_5$, a new round of optimizations based on $P'_1-P'_5$ is started until convergence of the parameter values is attained. The entire optimization requires typically 20 000−30 000 LTD minimizations.

After completing the optimization, an LTD run consisting of at least ~3000 minimized structures (with the optimal set of parameters) is carried out (in some cases longer runs up to 9000 structures were generated). These simulations always start from NOS, which is not a limitation as discussed earlier. The computer time required for the two components of the optimization procedure (i.e., LTD and minimizations of the partial group) depends on the size of the loop and the template. For example, an LTD run of 3000 minimizations of the (shortest) loop of acidic FGF (5 residues) and loop 2 of the antibody (8 residues and a large template) require ~70 and ~354 h CPU on an AMD Athlon 2.6 GHz processor, respectively. It should be pointed out that NOS1 undergoes further optimization during this procedure that might lead to a conformational change; this optimized NOS1 is denoted NOS2. Thus, NOS2 is used in the calculation of the final energy gaps, whereas the RMSD is

**TABLE 2: Optimized GB/SA Parameters for the Training Group of Loops and the Corresponding Energy Gaps[a]**

| | parameters | | | | | | energy gaps (kcal/mol) | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| protein/ loop | $P_1$ | $P_2$ | $P_3$ | $P_4$ | $P_5$ | $\sigma$ | Still's set[b] | $P_1-P_5$[c] | $P_1-P_5,\sigma$[d] | best-fit Still | $E_{FF}$ ($\epsilon = 2r$) eq 1 | optimal ASPs eq 1 | best-fit ASPs eq 1 |
| Still's set | 0.073 | 0.921 | 6.211 | 15.236 | 1.254 | 0.0049 | | | | | | | |
| best-fit set | −0.08 | 0.02 | 5.30 | 13.90 | 1.10 | 0.003 | | | | | | | |
| RNase A loop 3 89−97 (9) | 0.05 | −0.10 | 0.50 | 15.40 | 0.90 | 0.003 | 7.3<br>4.4 | **2.9** | **2.8** | 3.7 | 5.5 | **1.8** | **1.9** |
| RNase A, loop 1 64−71 (8) | −0.10 | 0.01 | 8.20 | 15.236 | 1.00 | 0.001 | 5.1<br>7.1 | **1.7** | **1.4** | **1.8** | **0.6** | | **0.4** |
| acidic FGF 90−94 (5) | 0.001 | 0.02 | 0.10 | 10.00 | 1.70 | 0.003 | 12.7<br>4.9 | 7.4 | 6.7 | 8.1 | 12.3 | 4.6 | 8.5 |
| adenylate kinase (AK) 73−80 (8) | −0.01 | 0.01 | −14.00 | −5.00 | 0.75 | 0.0045 | 11.1<br>14.3 | 7.4 | 7.2 | 9.2 | 11.8 | 6.0 | 13.5 |
| peptidase 205−213 (9) | 0.005 | 0.005 | 0.50 | 2.00 | 1.50 | −0.025 | **3.0**<br>4.6 | **1.2** | **0.7** | **1.3** | 4.1 | **0.5** | 6.0 |
| antibody loop 1 L26−37 (12) | −0.20 | 0.07 | 13.00 | 14.00 | 1.05 | −0.001 | 13.0<br>10.9 | 8.8 | 4.3 | 8.4 | 14.6 | 4.9 | 4.8 |
| antibody, loop 2 H102−109 (8) | −0.22 | 0.921 | 6.211 | 17.00 | 0.10 | 0.011 | 5.4<br>4.2 | **1.1** | **0.8** | **1.7** | 5.5 | **1.6** | **1.9** |
| penicillopepsin 129−137 (9) | −0.25 | −0.48 | 8.50 | 22.00 | 1.04 | 0.0005 | 6.1<br>6.5 | **3.0** | **1.6** | **2.6** | 10.3 | **1.8** | 4.1 |
| proteinase 202−210 (9) | 0.001 | −3.00 | 5.00 | 15.236 | 1.00 | 0.0034 | 3.2<br>3.4 | **2.7** | **2.6** | 5.0 | 4.9 | **0.5** | 3.4 |
| averages | | | | | | | 7.4 ± 1.3<br>6.7 ± 1.2 | 4.0 ± 1.0 | 3.1 ± 0.8 | 4.6 ± 3.2 | 7.7 ± 4.7. | 2.7 ± 0.7 | 4.9 ± 1.3 |

[a] Still's parameters $P_1-P_5$ and $\sigma$ are defined in eqs 3−8. Energy gaps between NOS2 and GEM were obtained by at least 3000 LTD minimizations. The energy gaps denoted $E_{FF}$, optimal ASPs, and best-fit ASPs are taken from paper II. Energy gaps smaller than 3 kcal/mol are bold-faced. The errors in the averages are one standard deviation divided by $n^{1/2}$ where $n = 9$. For the optimal ASPs, the result for loop 1 of RNase A (from paper II) is unavailable. Optimized ASPs results for loop 1 of RNase are not available. [b] Energy gaps obtained with Still's standard set of parameters, where the charge of Arg, Lys, Asp, Glu, and His is neutralized (upper row) and kept intact (lower row). [c] Energy gaps for optimized $P_1-P_5$. [d] Energy gaps for optimized $P_1-P_5$ and $\sigma$.

calculated with respect to NOS1. It is important to verify that NOS2 does not differ significantly from NOS1.

## Results and Discussion

**Optimization of the GB/SA Parameters and the Energy Gaps of the Training Group.** GB/SA is expected to model the electrostatic interactions better than eq 1; therefore, it was not clear a priori whether in the GB/SA parameter optimization the charges of Arg, Lys, His, Asp, and Glu should be neutralized as in paper II. To answer this question, we first applied Still's standard parameters ($P_1-P_5$, and $\sigma$) with charged and neutralized residues to the training group, that is, for each loop we carried out an LTD run of ~3000 minimizations [using $E_{tot}$ (eqs 3−5) where $E_{FF}$ is defined by the all-atom AMBER force field]. The corresponding energy gaps appear in Table 2 under "Still's set" where for each loop the results in the upper and lower rows are for the neutralized and charged residues, respectively. The table shows that overall the two sets of results are comparable with average gaps that are equal within the statistical errors. However, because for five out of the nine loops the neutralized set of results exhibit the lowest energy gaps, we decided to optimize the GB/SA parameters with neutralized charges on the loop and template. Notice also that according to our criterion both sets of gaps are too large because they exceed the 3 kcal/mol value, except for peptidase (neutralized). However, overall Still's results should be considered better than those obtained in paper II for $E_{FF}$ ($\epsilon = 2r$) (eq 1) that are provided as well. The $E_{FF}$ gaps from paper II are smaller than Still's neutralized and charged gaps only for three and two loops, respectively. Again, the average gap value obtained by $E_{FF}$ ($\epsilon = 2r$) does not provide a reliable measure of performance (even though it is slightly larger than those of Still's set) because of its large error bars, which reflect the strong scatter of the individual results. For most loops, the RMSD between NOS1 and NOS2 is small (less than 0.5 Å) except for proteinase and AK where the RMSD is 1.6 and 0.98 Å (for both the charged and neutralized loops), respectively. Therefore, the results for these loops should be evaluated with caution.

The table reveals that the optimized $P_1-P_5$ for the individual loops lead to a significant decrease in the energy gaps as compared to those obtained with Still's standard parameters and neutralized charge, and that with the optimizing both $P_1-P_5$ and $\sigma$ these values decrease further. Thus, for six of the loops, the gaps (bold-faced in the table) are smaller than 3 kcal/mol; correspondingly, the average gaps decrease significantly. However, for AK and proteinase the RMSD values between NOS1 and NOS2 are relatively large, 1.3 and 1.6 Å (for both optimal sets), respectively. The energy gaps obtained with the optimized ($P_1-P_5$) and the optimized ($P_1-P_5$ plus $\sigma$) are comparable to the energy gaps obtained with the optimized ASPs in paper II (see Table 2), which is also reflected by the average gap values. To reduce the gaps further, we attempted for several loops to optimize parameter $k$ (=4) of eq 3, and $\epsilon_{in}$, and $\epsilon_w$ of eq 4; however, we could not find parameter values that would lead to lower gaps.

The individual sets of optimized $P_1-P_5$ and $\sigma$ that appear in Table 2 constitute the basis for calculating the best-fit (bf) set. Although no definite prescription exists for such a derivation, a guiding principle would be to average the individual values, excluding parameters that deviate strongly from the others or reducing their absolute values. Thus, the best-fit $P_1$ and $P_5$ are exact and approximate averages over all the nine individual values, respectively. Best-fit $P_3$ and $P_4$ are averages over the individual values of eight loops, ignoring the strongly deviating values, −14.0 and −5.0 of AK, respectively. In the averages defining best-fit $P_2$ and $\sigma$ the *moderately* deviating values, −3.0

**TABLE 3: Results for the RMSD between NOS1 and the GEM Structure for the Training Group of Loops[a]**

| | RMSD (Å) | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Still's set | | | $P_1-P_5$ | | | $P_1-P_5$, $\sigma$ | | | best-fit Still | | | *E*FF ($\epsilon = 2r$) eq 1 | | | optimized ASPs, eq 1 | | | best-fit ASPs, eq 1 | | |
| protein/loop | BB | SC | TOT | BB | SC | TOT | BB | SC | TOT | BB | SC | TOT | BB | SC | TOT | BB | SC | TOT | BB | SC | TOT |
| RNase A, loop 3 | 0.5 | 0.5 | 0.5[b] | 0.6 | 1.7 | 1.3 | 0.5 | 1.3 | 1.0 | 0.8 | 2.1 | 1.6 | 0.5 | 1.5 | 1.1 | 0.3 | 1.2 | 0.9 | 0.2 | 1.4 | 1.0 |
| 89−97 (9) | 0.6 | 2.0 | 1.4 | | | | | | | | | | | | | | | | | | |
| RNase A, loop 1 | 0.7 | 1.9 | 1.4 | 0.5 | 2.1 | 1.4 | 0.5 | 2.2 | 1.5 | 0.5 | 2.0 | 1.3 | 0.4 | 0.9 | | 0.4 | 0.9 | | | | |
| 64−71 (8) | **1.4** | 3.1 | 2.3 | | | | | | | | | | | | | | | | | | |
| acidic FGF | 0.6 | 2.0 | 1.5 | 0.7 | 1.8 | 1.4 | 0.4 | 1.8 | 1.4 | **1.1** | 2.7 | 2.2 | 0.6 | 3.1 | 2.4 | 0.2 | 1.5 | 1.2 | 0.5 | 1.7 | 1.3 |
| 90−94 (5) | 0.9 | 2.4 | 1.9 | | | | | | | | | | | | | | | | | | |
| adenylate kinase | **1.2** | 4.1 | 2.9 | **1.1** | 3.5 | 2.5 | **1.1** | 3.8 | 2.7 | **2.8** | 7.2 | 5.3 | **1.1** | 3.5 | 2.5 | 1.0 | 3.2 | 2.3 | 0.9 | 2.9 | 2.1 |
| (AK) 73−80 (8) | 0.5 | 4.0 | 2.8 | | | | | | | | | | | | | | | | | | |
| peptidase | 0.1 | 1.0 | 0.7 | 0.1 | 1.5 | 1.1 | 0.4 | 1.7 | 1.3 | 0.1 | 1.3 | 0.9 | 0.8 | 1.3 | 1.0 | 0.2 | 0.9 | 0.6 | 0.2 | 1.2 | 0.9 |
| 205−213 (9) | 0.2 | 1.1 | 0.8 | | | | | | | | | | | | | | | | | | |
| antibody, loop 1 | **1.1** | 3.3 | 2.4 | **1.8** | 2.9 | 2.4 | **1.3** | 2.7 | 2.1 | **1.6** | 4.8 | 3.5 | **1.1** | 2.3 | 1.8 | 1.0 | 1.7 | 1.4 | 0.9 | 2.2 | 1.7 |
| L26−37 (12) | 1.0 | 1.5 | 1.2 | | | | | | | | | | | | | | | | | | |
| antibody, loop 2 | 1.0 | 3.2 | 2.5 | 0.3 | 0.6 | 0.5 | 0.4 | 1.4 | 1.1 | 1.0 | 3.1 | 2.5 | **1.2** | 0.9 | 1.0 | 0.7 | 0.7 | 0.7 | 0.6 | 0.9 | 0.8 |
| H102−109(8) | **1.1** | 1.3 | 1.2 | | | | | | | | | | | | | | | | | | |
| penicillopepsin | 0.4 | 1.9 | 1.3 | 0.2 | 0.8 | 0.6 | 0.2 | 1.1 | 0.8 | 0.3 | 0.8 | 0.6 | 0.4 | 1.9 | 1.4 | 0.1 | 1.2 | 0.9 | 0.1 | 1.5 | 1.0 |
| 129−137 (9) | 0.5 | 1.9 | 1.4 | | | | | | | | | | | | | | | | | | |
| proteinase | **1.4** | 1.7 | 1.5 | **1.4** | 1.7 | 1.5 | **1.4** | 1.7 | 1.5 | **1.4** | 1.7 | 1.5 | **1.9** | 3.1 | 2.4 | 0.3 | 1.2 | 0.7 | 0.4 | 1.3 | 0.8 |
| 202−210 (9) | **1.4** | 1.7 | 1.5 | | | | | | | | | | | | | | | | | | |
| averages | 0.8 | 2.2 | 1.6 | 0.7 | 1.8 | 1.4 | 0.7 | 2.0 | 1.5 | 1.1 | 2.9 | 2.2 | 0.9 | 2.2 | 1.6 | 0.5 | 1.5 | 1.1 | 0.5 | 1.6 | 1.2 |
| | 0.8 | 2.1 | 1.6 | | | | | | | | | | | | | | | | | | |
| SD/$n^{1/2}$ | 0.1 | 0.4 | 0.3 | 0.2 | 0.3 | 0.2 | 0.2 | 0.3 | 0.2 | 0.3 | 0.7 | 0.5 | 0.2 | 0.3 | 0.2 | 0.1 | 0.3 | 0.2 | 0.1 | 0.2 | 0.2 |
| | 0.1 | 0.3 | 0.2 | | | | | | | | | | | | | | | | | | |

[a] BB, SC, and TOT denote RMSD results for the backbone, side chains, and the total loop, respectively. The different columns are defined in the caption of Table 2. Backbone RMSD values larger than 1 Å are bold-faced. The corresponding errors in the averages appear in the bottom; SD is the standard deviation and $n = 9$. Some of the results for loop 1 of RNase A from paper II are unavailable. Not all of the results for loop 1 of RNase A are available. [b] RMSD results obtained with Still's standard set of parameters, where the charge of Arg, Lys, Asp, Glu, and His is neutralized (upper row) and kept intact (lower row).

of proteinase and −0.025 of peptidase were increased to −0.26 and −0.0002, respectively. Overall, the bf parameters are systematically lower than the corresponding Still's original values, where smaller $P_1$ leads to smaller $\alpha_i$ while smaller $P_2-P_4$ lead to larger $\alpha_i$ (eq 7).

It should be noted that for the bf parameters the RMSD values between NOS1 and NOS2 are all smaller than 0.85 Å (the value obtained for loop 1 of the antibody). The table shows that the energy gaps obtained with Still(bf) parameters are significantly better (lower) than the corresponding values based on Still's standard set for both neutralized and charged residues. There are two exceptions, namely, proteinase, where the values are 5 versus 3.2 kcal/mol, respectively, and acidic FGF (8.1 vs 4.9 kcal/mol) for charged residues. One must note, however, that the reliability of the results obtained for proteinase with Still's standard parameters is somewhat questionable because of the large RMSD between NOS1 and NOS2 mentioned above. Also, the energy gaps for Still's(bf) are slightly better than those obtained by ASPs(bf), where four and three gaps are smaller than 3 kcal/mol, respectively (the average gaps are comparable).

**RMSD for the Training Group.** The RMSD between the GEM structure and NOS1 is calculated with respect to the heavy atoms and without superposition on NOS1 (the same applies to RMSD between NOS1 and NOS2 discussed earlier). An accepted criterion for a successful prediction of the loop backbone (BB) structure is that the RMSD from the correct structure is not larger than 1 Å;[34,35] notice, however, that RMSD values smaller than 0.4 Å are actually insignificant because the two structures belong to the same microstate.

RMSD results (between NOS1 and GEM) for the training set of loops are summarized in Table 3, which is structured similar to Table 2. In particular, two sets of results are presented in the column "standard Still" where for each loop the first and second row contains results obtained with neutralized and charges residues, respectively. The RMSD values are given for the backbone (BB), the side chains (SC), and the total loop (TOT). The general observation is that for all methods and optimizations the BB results are quite satisfactory. Thus, for each of Still's standard sets (i.e., charged and neutralized), only three RMSD values (bold-faced in the table) are larger than 1 Å, where they do not exceed 1.4 Å. The same tendency with minor changes characterizes all Still's results, where the largest RMSD(BB) values occur for proteinase with 1.4 Å for all approximations and loop 1 of antibody and AD with maximal values of 1.8 Å ($P_1-P_5$) and 2.8 Å (bf), respectively. It is evident that Still's (bf) results are slightly inferior to the other sets of Still(BB) values but they are comparable to results based on the force field alone [$E_{FF}$ ($\epsilon = 2r$)], where four deviations larger than 1 Å also occur. However, the RMSD(BB) values for the optimized ASPs and ASPs(bf) are *all* within the range of 1 Å and thus are better than any of Still's sets; these trends are also reflected by the averages of the optimized ASPs and ASPs(bf) that are slightly lower than the other averages.

Most of the RMSD(SC) results are larger than 1 Å, and for standard Still the charged and neutral results are almost comparable (for four out of seven loops the neutral RMSD-(SC) results are smaller than the charged values while the averages are actually identical). The RMSD(SC) results for the optimized $P_1-P_5$ and optimized $P_1-P_5$ plus $\sigma$ are comparable and are slightly better (for five out of eight loops) than the standard Still values (neutral and charged). As is shown clearly in the table, Still(bf) results for RMSD(SC) are inferior to those of the other Still's approximations and even to those obtained by the force field [$E_{FF}$ ($\epsilon = 2r$)]; this is also reflected by the relatively high average of 2.9 Å for Still(bf). The best results are obtained for the optimized ASPs and ASPs(bf), where the average RMSD(SC) values are 1.5 and 1.6 Å, respectively; however, notice that within the error bars these values are equal

**TABLE 4: Energy Gaps for the Test Group of Loops[a]**

| | energy gaps (kcal/mol) | | | |
|---|---|---|---|---|
| protein, loop | standard Still[b] | best-fit Still | $E_{FF}$ ($\epsilon = 2r$) eq 1 | best-fit ASPs eq 1 |
| ser-proteinase | 17.6 | 13.7 | 6.9 | 3.9 |
| 143−151(9) | 12.0 | | | |
| proteinase | 7.4 | 9.3 | 10.0 | 4.7 |
| 188−196 (9) | 1.8 | | | |
| proteinase | 25.4 | 13.4 | 14.8 | 3.3 |
| 128−137 (10) | 10.6 | | | |
| peptidase | 9.9 | 6.1 | 9.0 | 3.4 |
| 244−250 (7) | 9.7 | | | |
| RNase H | 11.9 | 7.5 | 14.0 | 9.4 |
| 57−63 (7) | 11.7 | | | |
| antibody | 8.0 | 6.6 | 9.8 | 8.8 |
| 56L−62L (7) | 5.6 | | | |
| antibacterial pro. | 0.0 | 0.0 | 0.5 | 1.7 |
| 25−30 (6) | 1.2 | | | |
| averages | $11.5 \pm 3.1$ | $8.1 \pm 1.8$ | $9.3 \pm 1.8$ | $5.0 \pm 1.1$ |
| | $7.5 \pm 1.7$ | | | |

[a] Energy gaps between NOS2 and the GEM were obtained by at least 3000 LTD minimizations. The energy gaps denoted $E_{FF}$ and the best-fit ASPs are taken from paper II. The errors in the averages are one standard deviation divided by $n^{1/2}$ where $n = 7$. [b] Energy gaps obtained with Still's standard set of parameters, where the charge of Arg, Lys, Asp, Glu, and His is neutralized (upper row) and kept intact (lower row).

to those obtained for Still's set, with optimized $P_1-P_5$, and optimized $P_1-P_5$ plus $\sigma$.

**Energy Gaps for the Test Group.** The energy gaps obtained by various methods for a test group of 7 loops are summarized in Table 4. As in Tables 2 and 3, for each loop results presented in the upper and lower rows of the second column were calculated with Still's standard parameters with neutralized and charged amino acids, respectively; we start by discussing these results. It should first be noted that the *R* values of the last four loops are relatively small (1.3−2.7; see Table 1), suggesting that these loops are only moderately flexible. This is probably reflected in the comparable energy gaps obtained for each pair, even though the loops of RNase H and antibody consist of a relatively large number of charged amino acid residues, that is, 3 and 2, respectively (as pointed out earlier, even after charge neutralization these residues still have significant dipole moments). Notice also that for the last loop (of antibacterial protein) the gap obtained with standard Still(neutralized) is zero, meaning that the GEM structure = NOS2, where for Still(charged) this gap is small, 1.2 kcal/mol. All of these results are reliable in the sense that for each loop the RMSD between NOS1 and NOS2 is smaller than 0.56 Å obtained for RNase H.

The first three loops in Table 4 are the longest (9, 9, and 10 residues), are characterized by relatively large *R* values (4.9, 4.5, and 4.3, see Table 1), and they contain one, two, and three charged residues, respectively. The energy gaps obtained for these loops with Still's standard parameters and charged residues are always significantly smaller than those obtained with the neutralized charge. While such large differences are not unexpected for these potentially flexible loops, part of these results might not be reliable because of large RMSD values between NOS1 and NOS2. For ser-proteinase these RMSD values are small, 0.27 and 0.23 Å for the neutralized and charged residues, respectively; however, they are large for loop 188−196 of proteinase (1.51 and 0.86 Å, respectively), and very large (2.39 Å) for loop 128−137 of proteinase (charged). In this respect, Still's bf gaps are more reliable because the RMSD values between NOS1 and NOS2 are smaller than 0.66 Å. As expected, the bf energy gaps are smaller than their counterparts obtained with Still's standard parameters and neutralized charges, except for loop 188−196 of proteinase where the reliability of 7.4 kcal/mol obtained with Still's standard parameters is questionable, as discussed above.

**TABLE 5: Results for the RMSD between NOS1 and the GEM Structure for the Test Group of Loops[a]**

| | RMSD (Å) | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | standard Still | | | best-fit Still | | | $E_{FF}$ ($\epsilon = 2r$) eq 1 | | best-fit ASP's eq 1 | |
| protein/loop | BB | SC | TOT | BB | SC | TOT | BB | TOT | BB | TOT |
| ser-proteinase | 0.2 | 1.0 | 0.7[b] | 0.2 | 1.7 | 1.1 | **2.1** | 2.4 | 0.6 | 0.6 |
| 143−151 (9) | 0.2 | 0.9 | 0.6 | | | | | | | |
| proteinase | 0.7 | 2.3 | 1.7 | 0.7 | 2.3 | 1.8 | 0.3 | 1.5 | 0.2 | 0.9 |
| 188−196 (9) | 0.3 | 1.9 | 1.4 | | | | | | | |
| proteinase | **1.1** | 2.8 | 2.1 | **1.1** | 4.8 | 3.4 | **1.3** | 2.3 | 0.8 | 1.0 |
| 128−137 (10) | **1.5** | 2.9 | 2.4 | | | | | | | |
| peptidase | 0.7 | 1.5 | 1.2 | 0.7 | 1.6 | 1.3 | 0.7 | 1.3 | 0.6 | 2.2 |
| 244−250 (7) | 0.8 | 2.0 | 1.5 | | | | | | | |
| RNase H | **1.5** | 3.7 | 2.8 | 0.8 | 2.8 | 2.1 | 0.2 | 1.5 | 0.2 | 1.9 |
| 57−63 (7) | 0.9 | 1.9 | 1.5 | | | | | | | |
| antibody | 0.8 | 1.4 | 1.1 | 0.8 | 1.3 | 1.0 | 0.1 | 0.7 | 0.7 | 1.1 |
| 56L-62L (7) | 0.8 | 0.9 | 0.8 | | | | | | | |
| antibacterial | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.1 | 0.6 | 0.1 | 0.7 |
| prot. 25−30 (6) | 0.2 | 0.2 | 0.2 | | | | | | | |
| averages | 0.7 | 1.8 | 1.4 | 0.6 | 2.1 | 1.5 | 0.7 | 1.5 | 0.5 | 1.2 |
| | 0.7 | 1.5 | 1.2 | | | | | | | |
| SD/$n^{1/2}$ | 0.2 | 0.5 | 0.4 | 0.1 | 0.6 | 0.4 | 0.3 | 0.3 | 0.1 | 0.2 |
| | 0.2 | 0.3 | 0.3 | | | | | | | |

[a] BB, SC, and TOT denote RMSD results for the backbone, side chains, and the total loop, respectively. The different columns are defined in the caption of Table 2. Backbone RMSD values larger than 1 Å are bold-faced. The corresponding errors in the averages appear in the bottom; SD is the standard deviation and $n = 7$. In paper II, the results for SC are not provided. [b] RMSD results obtained with Still's standard set of parameters, where the charge of Arg, Lys, Asp, Glu, and His is neutralized (upper row) and kept intact (lower row).

The gaps obtained with Still's best-fit parameters are also smaller than those obtained by the force field alone [$E_{FF}$ ($\epsilon = 2r$)] in paper II, which are also presented in the Table; the only exception occurs for ser-proteinase. Alternatively, for the first four loops the gaps obtained with ASPs(bf) in paper II are significantly smaller than those obtained with Still(bf), while for the last three loops Still(bf)'s gaps are slightly smaller. This again demonstrates that the simplified model (eq 1) is better than the more sophisticated GB/SA model. This is also demonstrated by the average value for ASPs(bf), $5.1 \pm 1.1$ kcal/mol that is smaller than most of the other averages in the table, where it is only equal (within the error bars) to $7.5 \pm 1.7$ kcal/mol obtained for Still's standard parameters (charged).

**RMSD for the Test Group.** RMSD results for the test group appear in Table 5, and as for the training group, we discuss them first for the backbone [RMSD(BB)]. For Still's standard parameters most of the RMSD are smaller than 1 Å besides RMSD = 1.5 Å obtained for loop 128−137 of proteinase (charged residues). A relatively large value, 1.5 Å, is also shown for RNAse H (neutralized), where this value decreases to 0.8 Å for Still(bf); the other RMSD(BB) results remain the same for Still(standard) and Still(bf). The RMSD(BB) values for the force field alone [$E_{FF}$ ($\epsilon = 2r$)] are larger than those of Still(bf) for ser-proteinase (2.1 vs 0.2 Å) and for loop 128−137 of proteinase (1.3 vs 1.1 Å); for the rest of the loops the force field results are predominantly the lowest and they are smaller than 1 Å. However, the lowest set of RMSD(BB) is again that of ASPs(bf) where all are smaller than 1 Å. However, all of the averages are below 1 Å and they are equal within the error bars.

**TABLE 6: Average Energy Gaps and RMSD Values for the 16 Loops of the Training and Test Groups[a]**

| | standard Still (neutralized)[b] | standard Still (charged)[c] | best-fit Still | *E*FF ($\epsilon = 2r$), eq 1 | best- fit ASP's |
|---|---|---|---|---|---|
| | | | average energy gaps (kcal/mol) | | |
| | 9.75 ± 1.73 | 7.06 ± 1.08 | 6.15 ± 1.04 | 8.40 ± 1.15 | 5.00 ± 0.87 |
| | | | average RMSD (Å) | | |
| BB[d] | 0.75 ± 0.11 | 0.77 ± 0.11 | 0.87 ± 0.17 | 0.80 ± 0.15 | 0.46 ± 0.07 |
| SC | 2.02 ± 0.30 | 1.86 ± 0.23 | 2.51 ± 0.45 | | |
| TOT | 1.51 ± 0.21 | 1.43 ± 0.17 | 1.88 ± 0.33 | 1.55 ± 0.17 | 1.18 ± 0.13 |

[a] The errors in the averages are one standard deviation divided by $n^{1/2}$ where $n = 16$. [b] Calculated with Still's standard parameters with neutralized charge for Arg, Lys, Glu, Asp, and His. [c] Calculated with Still's standard parameters with charged residues. [d] BB = backbone; SC = side chains; TOT = total.

The RMSD(SC) results obtained for Still's standard parameters, as expected, are larger than the corresponding RMSD(BB) values and in most cases are larger than 1 Å. However, these values (for the neutral residues) are not worse (and in three cases they are actually better) than the corresponding values obtained for Still(bf); the same applies to the total RMSD values [RMSD(TOT)]. In paper II results were presented for RMSD(TOT) but not for RMSD(SC), which therefore do not appear in Table 5. The table reveals that in five out of seven cases the RMSD(TOT) values obtained with the force field alone or with ASPs(bf) are equal or smaller (better) than those of Still(bf). For Still(bf) the largest RMSD(TOT) is 3.4 Å (proteinase, 128−137), where the largest values obtained with the force field and ASPs(bf) are smaller, 2.4 Å (ser-proteinase), and 2.2 Å (peptidase), respectively. Notice that for five loops the ASPs(bf) TOT values are not larger than 1.1 Å! The averages of RMSD(TOT) follow the above trends but statistically they are all equal.

**Overall Evaluation of the Different Models.** The above discussion of results already demonstrates some advantage of eq 1 over the GB/SA model. To evaluate these models further, we present in Table 6 averages calculated over the entire group of 16 loops for the energy gaps and the RMSD values as well as their standard deviations (divided by $16^{1/2} = 4$). As expected, for the three Still's models, the lowest average energy (6.15 kcal/mol) is obtained with the bf parameters; this value is significantly smaller (i.e., beyond the statistical errors) than 9.75 obtained by Still's original parameters with neutralized residues and 8.4 kcal/mol obtained by the force field alone [$E_{FF}$ ($\epsilon = 2r$), eq 1]. However, 6.15 is equal within the statistical errors to the slightly larger gap, 7.06 kcal/mol obtained for Still's original parameters with charged residues. The lowest gap, 5.0 kcal/mol (with the lowest statistical error) is observed for ASPs(bf); however, within the error bars, this value should be considered equal to 6.15. Correspondingly, the backbone RMSD of ASPs(bf), 0.46 Å, is significantly lower than the values obtained with the other models, where the latter results are equal within the error bars. Also, the RMSD(TOT) result, 1.18 Å for ASPs(bf) is the lowest; however, its error overlaps those of Still's(standard).

Thus, although the advantage of eq 1 with ASPs(bf) over Still's results is in most cases statistically significant, the distinction between the performance of Still's models would require results from a larger sample of loops. However, the trend shown in the table is that Still(bf) provides the lowest average energy gap (among Still's models) while its RMSD values are somewhat inferior to those of the other models. In retrospect, the fact that comparable results obtained for Still's models is perhaps not surprising because the standard and best-fit sets of parameters are in most cases not very different, where the four best examples are $P_3$, $P_4$, $P_5$, and $\sigma$ that are 6.211 versus 5.30, 15.236 versus 13.90, 1.254 versus 1.10, and 0.0049 versus 0.0030 for Still(standard) and Still(bf), respectively (see Table 2). This should be compared to the more drastic changes occurred in the optimization of the ASPs in paper II, where the optimized (and bf) value of carbon (which is the most frequent atom) has been found to be negative (hydrophilic) versus its positive value (hydrophobic) in the sets of Wesson and Eisenberg,[49] and Ooi et al.,[77] for example. This may suggest that the original (standard) optimization of Still's parameters against PB results for small molecules is reasonable, a fact that could not have been gathered a priori. However, our hope that GB/SA would provide better results than the theoretically inferior eq 1 has not been materialized to our surprise (and disappointment); the reason for this unexpected behavior remains unclear. Still, it is possible that other GB/SA versions would provide better results for loops than the present model.

**Attempts to Improve Equation 1.** In view of the above discussion, it would be of interest to check whether eq 1 can still be improved. As has already been pointed out and discussed in more detail in paper II, the dielectric function, $\epsilon = nr$ with $n = 2$ used for optimizing the ASPs does not provide the necessary screening of the Coulombic interactions for a loop consisting of several charged residues (even if neutralized), while increasing the screening to $\epsilon = 3r$ made eq 1 insensitive to conformational changes and thus did not allow optimization of the ASPs. To overcome this problem, we decided to replace the $\epsilon = nr$ function by more complex dielectric functions and study their performance. The first function, used by Mehler and collaborators is[102,103]

$$\epsilon(r) = (\epsilon_w + 1)/(1 + k\exp[-\lambda(\epsilon_w - 1)r]) - 1 \quad (10)$$

where $\epsilon_w = 80$, and $k$ and $\lambda$ are parameters to be optimized. The second function, proposed by Warshel is[104]

$$\epsilon(r) = \begin{cases} 16.55 & r < 3\text{Å} \\ 1 + 60(1 - \exp(-0.1r)) & r \geq 3\text{Å} \end{cases} \quad (11)$$

where both functions have been implemented within TINKER. Equation 10 was applied to loop 3 of RNase A and the loop of acidic fibroblast, where both $\epsilon_0$ and $\lambda$, and the ASPs were optimized. Equation 11 was applied to loop 3 of RNase A and the second loop of proteinase (of the test group). Here no parameters exist and thus only the ASPs were optimized. However, in both cases we could not obtain better energy gaps than those obtained with $\epsilon = 2r$.

**Other Recent Studies of Loops.** Still's GB/SA model with the AMBER force field has been applied recently to loops by de Bakker et al.[52,53] who treated 385 loop targets (length 2 to 12) collected previously by Fiser et al.[44] For each target, a set of 1000 decoy structures were generated using the RAPPER and SCRWL search procedures for the backbone and side chains, respectively. The energies of these decoys were than

minimized with the GB/SA/AMBER function and for comparison also by the AMBER force field (with $\epsilon = 1$) alone, using the program TINKER. As in our studies, they have found in general a better performance with GB/SA/AMBER than with AMBER alone. Later, an extensive study of loops was carried out by Jacobson et al.[54] who used the surface generalized Born and a nonpolar solvation model (SGB-NP)[56] with the latest version of the OPLS force field.[102] They have treated a full set of 788 target loops (length 4−12) and a filtered set of 514 loops, where for each loop 200−1400 decoys have been generated by an elaborate conformational search procedure. Very recently, Zhang et al.[58] have tested their knowledge-based statistical potential, DFIRE (distance-scaled, finite ideal gas reference state) by applying it to these three loop sets and comparing its performance to those of GB/SA/AMBER and SGB-NP/OPLS. From these results, one can obtain some information about the relative performance of the above models.

Thus, in the section "Minimized" of Table S2 of the supplemental material provided by Zhang et al.[58] the average RMSD results obtained by GB/SA/AMBER and DFIRE for different loop length are presented. Dividing the provided standard deviation values by $n^{1/2}$ where $n$ is the number of loops of certain length studied, show that only for three loop sizes, 3, 4, and 6, the values of GB/SA/AMBER are smaller than those of DFIRE, whereas in all other cases the corresponding results are equal within the error bars. However, in the section "Full" of Table S4 OPLS/SGB-NP leads to smaller RMSD values than DFIRE for six loop lengths (from 4 to 9), where for the longer loops (10−12) the results are equal within the statistical errors. For the filtered set, OPLS/SGB-NP leads to the smallest RMSD values for five loop lengths (4−8) where for the longer loops (9−12) the results are equal results within the statistical errors.

Thus, OPLS/SGB-NP performs better with respect to DFIRE than does AMBER/GB/SA, suggesting that OPLS/SGB-NP is the more reliable model among the two at least for loops. Clearly, this conclusion should be taken with some caution because the RAPPER set is smaller and different from Jacobson's sets, and from our experience, the number of decoys used in these studies is insufficient. In our studies, for example, 3000−9000 conformations are generated for each loop in a search process (LTD) that directs the loop toward its GEM structure. Also, it is not clear what is the relative contribution of the force fields to the performance of these models. In paper I we have found AMBER to be better than OPLS for loops but the torsional potentials of OPLS have been recently improved[105] and used in the OPLS/SGB-NP study.

This discussion is closely related to recent performance studies of GB/SA solvation models. It has been found that some combinations of force fields and GB/SA models are better than others and can lead to results that are close to those obtained in the experiment or by explicit solvation models. A well-studied example is the (caped) C-terminal polypeptide from the B1 domain of protein G, a 16-residue peptide that has been found experimentally to fold to a $\beta$ hairpin in aqueous solutions.[106−108] Folding simulations based on different *explicit* water models (TIP3P, SPC) and force fields have all found the $\beta$-hairpin state to be the most populated.[109−112] However, simulations of Zhou and Berne,[113] Zhou,[112] and Levy's group[114] have shown that only few of the implicit models studied predict the $\beta$-hairpin state to be the most stable.

## Conclusions

All of the solvation models studied here [including $E_{FF}$ ($\epsilon = 2r$)] are considerably better than those using the force field with $\epsilon = 1$ [ $E_{FF}$ ($\epsilon = 1$)] as has been discussed in papers I and II. On the basis of results for 16 loops, we have not found significant differences in performance among the three GB/SA models studied. All of them, however, have been shown to be somewhat inferior to eq 1, which itself is unsatisfactory, leading to too high energy gaps of ∼5 kcal/mol. We have also concluded (indirectly) about differences in the performance of DFIRE[58] and the models of de Bakker et al.[52,53] and Jacobsen et al. However, these differences (based on the average behavior) are not very large as well, and for certain individual loops are reversed. It should be pointed out that for loops shorter than eight residues RMSD(BB) obtained by all these models is satisfactory.

Implicit solvation models are very convenient for studying loops because of their relative simplicity and the fact that they are amenable to efficient conformational search techniques. The problem is whether they can be improved significantly further. In this context, it should be emphasized again that most of the loop studies (excluding DFIRE) are based on minimized energy structures, where RMSD differences of 0.1−0.5 Å are insignificant because the corresponding structures belong to the same microstate. Neglecting the conformational entropy also hampers the search for correlation between RMSD and the free energy gap. Preliminary calculations in paper II have shown, however, that the contribution of the entropy has led to an insufficient decrease in the free energy gaps, that is, only by ∼0.6 kcal/mol. Entropic effects have been included successfully in the colony free energy.[59,115] Better agreement with the experimental data can expected to be achieved by taking into account the crystal environment and the effect of ions, and by selecting loops with low B factors.[54,116]

An important factor that affects the quality of loop modeling is an optimal match between a given implicit solvation model and the force field used. To be consistent with papers I and II, we have applied here GB/SA with AMBER94; however, extensive studies of the C-terminal polypeptide from the B1 domain of protein G by Zhou using AMBERx/GBSA,[109] where $x = 94$, 96, and 99 discovered that only AMBER96 (ref 117) with GB/SA gave a reasonable free energy profile (but one erroneous salt bridge); therefore, optimizing eq 1 with AMBER96 or other new optimized force fields might have improved this model further. One perhaps might choose GB models that maximally mimic of the Poisson−Boltzmann equation; however, Lee and co-workers[118,119] have argued recently that PB itself has its limitation and one has to resort to explicit−implicit hybrid models. Thus, developing the optimal implicit solvation model in general and for loops in particular still remains an open problem.[120]

**Acknowledgment.** This work was supported by NIH grants R01GM61916 and R01GM66090 and by National Science Foundation Large Information Technology Research Grant NSF0225636.

## References and Notes

(1) Karplus, P.A.; Schulz, G. E. *Naturwissenschaften* **1985**, *72*, 212.
(2) Getzoff, E. D.; Geysen, H. M.; Rodda, S. J.; Alexander, H.; Tainer, J. A.; Lerner, R. A. *Science* **1987**, *235*, 1191.
(3) Rini, J. M.; Schulze-Gahmen, U.; Wilson, I. A. *Science* **1992**, *255*, 959.
(4) Constantine, K. L.; Friedrichs, M. S.; Wittekind, M.; Jamil, H.; Chu, C. H.; Parker, R. A.; Goldfarb, V.; Mueller, L.; Farmer, B. T. *Biochemistry* **1998**, *37*, 7965.
(5) Nicholson, L. K.; Yamazaki, T.; Torchia, D. A.; Grzesiek, S.; Bax, A.; Stahl, S. J.; Kaufman, J. D.; Wingfield, P. T.; Lam, P. Y. S.; Jadhav, P. K.; Hodge, C. N.; Domaille, P. J.; Chang, C.-H. *Struct. Biol.* **1995**, *2*, 274.
(6) Collins, J. R.; Burt, S. K.; Erickson, J. W. *Struct. Biol.* **1995**, *2*, 334.
(7) Wagner, G. *Struct. Biol.* **1995**, *2*, 255.
(8) Fetrow, J. S. *FASEB J.* **1995**, *9*, 708.

(9) Bates, P. A.; Sternberg, M. J. *Proteins* **1999**, *Suppl 3*, 47.
(10) Mosimann, S.; Meleshko, R.; James, M. N. *Proteins* **1995**, *23*, 301.
(11) Sali, A. *Curr. Opin. Biotechnol.* **1995**, *6*, 437.
(12) Petrey, D.; Xiang, Z.; Tang, C. L.; Xie, L.; Gimpepev, M.; Mitros, T.; Soto, C. S.; Goldsmith-Fischman, S.; Kernytsky, A.; Schlessinger, A.; Koh, I. Y. Y.; Alexov, E.; Honig, B. *Proteins* **2003**, *53*, 430.
(13) Crasto, C. J.; Feng, J. *Proteins* **2001**, *42*, 399.
(14) Leszczynski, J. F.; Rose, G. D. *Science* **1986**, *234*, 849.
(15) Donate, L. E.; Rufino, S. D.; Canard, L. H.; Blundell, T. L. *Protein Sci.* **1996**, *5*, 2600.
(16) Fechteler, T.; Dengler, U.; Schomburg, D. *J. Mol. Biol.* **1995**, *253*, 114.
(17) Kwasigroch, J. M.; Chomilier, J.; Mornon, J. P. *J. Mol. Biol.* **1996**, *259*, 855.
(18) Martin, A. C.; Toda, K.; Stirk, H. J.; Thornton, J. M. *Protein Eng.* **1995**, *8*, 1093.
(19) Oliva, B.; Bates, P. A.; Querol, E.; Aviles, F. X.; Sternberg, M. J. *J. Mol. Biol.* **1997**, *266*, 814.
(20) Ring, C. S.; Kneller, D. G.; Langridge, R.; Cohen, F. E. *J. Mol. Biol.* **1992**, *224*, 685.
(21) Pal, M.; Dasgupta, S. *Proteins* **2003**, *51*, 591.
(22) Chothia, C.; Lesk, A. M. *J. Mol. Biol.* **1987**, *196*, 901.
(23) Chothia, C.; Lesk, A. M.; Tramontano, A.; Levitt, M.; Smith-Gill, S. J.; Air, G.; Sheriff, S.; Padlan, E. A.; Davies, D.; Tulip, W. R. *Nature* **1989**, *342*, 877.
(24) Fidelis, K.; Stern, P. S.; Bacon, D.; Moult, J. *Protein Eng.* **1994**, *7*, 953.
(25) Espandaler, J.; Fernandez-Fuentes, N.; Hermoso, A.; Querol, E.; Aviles, F. X.; Sternberg, M. J. E.; Oliva, B. *Nucleic Acids Res.* **2004**, *32*, D185.
(26) Summers, N. L.; Karplus, M. *J. Mol. Biol.* **1990**, *216*, 991.
(27) Tappura, K. *Proteins* **2001**, *44*, 167.
(28) Wohlfahrt, G.; Hangoc, V.; Schomburg, D. *Proteins* **2002**, *47*, 370.
(29) Deane, C. M.; Blundell, T. L. *Proteins* **2000**, *40*, 135.
(30) van Vlijmen, H. W.; Karplus, M. *J. Mol. Biol.* **1997**, *267*, 975.
(31) Wojcik, J.; Mornon, J. P.; Chomilier, J. *J. Mol. Biol.* **1999**, *289*, 1469.
(32) Samudrala, R.; Moult, J. *J. Mol. Biol.* **1998**, *275*, 895.
(33) Sudarsanam, S.; DuBose, R. F.; March, C. J.; Srinivasan, S. *Protein Sci.* **1995**, *4*, 1412.
(34) Bruccoleri, R. E.; Karplus, M. *Biopolymers* **1987**, *26*, 137.
(35) Moult, J.; James, M. N. *Proteins* **1986**, *1*, 146.
(36) Fine, R. M.; Wang, H.; Shenkin, P. S.; Yarmush, D. L.; Levinthal, C. *Proteins* **1986**, *1*, 342.
(37) Higo, J.; Collura, V.; Garnier, J. *Biopolymers* **1992**, *32*, 33.
(38) Rosenfeld, R.; Zheng, Q.; Vajda, S.; DeLisi, C. *J. Mol. Biol.* **1993**, *234*, 515.
(39) Shenkin, P. S.; Yarmush, D. L.; Fine, R. M.; Wang, H. J.; Levinthal, C. *Biopolymers* **1987**, *26*, 2053.
(40) Caralacci, L.; Englander, S. W. *J. Comput. Chem.* **1996**, *17*, 1002.
(41) Dudek, M. J.; Scheraga, H. A. *J. Comput. Chem.* **1990**, *11*, 121.
(42) Gō, N.; Scheraga, H. A. *Macromolecules* **1970**, *3*, 178.
(43) Zheng, Q.; Rosenfeld, R.; Vajda, S.; DeLisi, C. *J. Comput. Chem.* **1993**, *14*, 556.
(44) Fiser, A.; Do, R. K. G.; Šali, A. *Protein Sci.* **2000**, *9*, 1753.
(45) Das, B.; Meirovitch, H. *Proteins* **2001**, *43*, 303.
(46) Das, B.; Meirovitch, H. *Proteins* **2003**, *43*, 470.
(47) Mas, M. T.; Smith, K. C.; Yarmush, D. L.; Aisaka, K.; Fine, R. M.; *Proteins* **1992**, *14*, 483.
(48) Smith, K. C.; Honig, B. *Proteins* **1994**, *18*, 119.
(49) Wesson, L.; Eisenberg, D. *Protein Sci.* **1992**, *1*, 227.
(50) Qiu, D.; Shenkin, P. S.; Hollinger, F. P.; Still, W. C. *J. Phys. Chem A* **1997**, *101*, 3005.
(51) Rapp, C. S.; Friesner, R. A. *Proteins* **1999**, *35*, 173.
(52) de Bakker, P. I. W.; DePristo, M. A.; Burke, D. F.; Blundell, T. L. *Proteins* **2003**, *51*, 21.
(53) DePristo, M. A.; de Bakker, P. I. W.; Lovell, S. C.; Blundell, T. L. *Proteins* **2003**, *51*, 41.
(54) Jacobson, M. P.; Pincus, D. L.; Rapp, C. S.; Day, T. J. F.; Honig, B.; Shaw, D. E.; Friesner, R. A. *Proteins* **2004**, *55*, 351.
(55) Ghosh, A.; Sendrovic Rapp, C.; Friesner, R. *J. Phys. Chem. B* **1998**, *102*, 10983.
(56) Gallicchio, E.; Zhang, L. Y.; Levy, R. M. *J. Comput. Chem.* **2002**, *23*, 517.
(57) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225.
(58) Zhang, C.; Liu, S.; Zhou, Y. *Protein Sci.* **2004**, *13*, 391.
(59) Xiang, X.; Soto, C. S.; Honig, B. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 7432.
(60) Baysal, C.; Meirovitch, H. *J. Am. Chem. Soc.* **1998**, *120*, 800.
(61) Baysal, C.; Meirovitch, H. *Biopolymers* **1999**, *50*, 329.
(62) Meirovitch, H.; Meirovitch, E.; Lee, J. *J. Phys. Chem.* **1995**, *99*, 4847.
(63) Meirovitch, H.; Meirovitch, E. *J. Phys. Chem.* **1996**, *100*, 5123.
(64) Baysal, C.; Meirovitch, H. *Biopolymers* **2000**, *54*, 416.
(65) Baysal, C.; Meirovitch, H. *Biopolymers* **2000**, *53*, 423.
(66) Maiorov, V. N.; Crippen, G. M. *J. Mol. Biol.* **1992**, *227*, 876.
(67) Mirny, L. A.; Shakhnovich, E. I. *J. Mol. Biol.* **1996**, *264*, 1164.
(68) Seok, C.; Rosen, J. B.; Chodera, J. D.; Dill, K. A. *J. Comput. Chem.* **2003**, *24*, 89.
(69) Baysal, C.; Meirovitch, H. *J. Phys. Chem.* **1997**, *101*, 2185.
(70) Baysal, C.; Meirovitch, H. *J. Comput. Chem.* **1999**, *20*, 1659.
(71) Meirovitch, H. *Chem. Phys. Lett.* **1977**, *45*, 389.
(72) Meirovitch, H.; Koerber, S. C.; Rivier, J. E.; Hagler, A. T. *Biopolymers* **1994**, *34*, 815.
(73) White, R. P.; Meirovitch, H. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 9235.
(74) Cheluvaraja, S.; Meirovitch, H. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 9241.
(75) Tanner, J. J.; Nell, L. J.; McCammon, J. A. *Biopolymers* **1992**, *32*, 23.
(76) Lins, R. D.; Briggs, J. M.; Straatsma, T. P.; Carlson, H. A.; Greenwald, J.; Choe, S.; McCammon, J. A. *Biophys. J.* **1999**, *76*, 2999.
(77) Ooi, T.; Oobatake, M.; Nemethy, G.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A* **1987**, *84*, 3086.
(78) Schiffer, C. A.; Caldwell, J. W.; Kollman, P. A.; Stroud, R. M. *Mol. Simul.* **1993**, *10*, 121.
(79) Fraternali, F.; Van Gunsteren, W. F. *J. Mol. Biol.* **1996**, *256*, 939.
(80) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M., Jr.; Ferguson, D. M.; Spellmeyer, D, C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179.
(81) Ponder, J. W. *TINKER-Software Tools for Molecular Design*, version 3.9; Washington University, St. Louis, MO, 2001.
(82) Najmanovich, R.; Kuttner, J.; Sobolev, V.; Edelman, M. *Proteins* **2000**, *39*, 261.
(83) Zhao, S.; Goodsell, D. S.; Olson, A. J. *Proteins* **2001**, *43*, 271.
(84) Wilson, M. A.; Brunger, A. T. *J. Mol. Biol.* **2000**, *301*, 1237.
(85) Esposito, L.; Vitagliano, L.; Sica, F.; Sorrentino, G.; Zagari, A.; Mazzarella, L. *J. Mol. Biol.* **2000**, *297*, 713.
(86) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. *J. Am. Chem. Soc.* **1990**, *112*, 6127.
(87) Hawkins, G. D.; Liotard, D. A.; Cramer, C. J.; Truhlar, D. G. *J. Org. Chem.* **1998**, *63*, 4305.
(88) Schaefer, M.; Karplus, M. *J. Phys. Chem.* **1996**, *100*, 1578.
(89) Jayaram, B.; Liu, Y.; Beveridge, D. L. *J. Chem. Phys.* **1998**, *109*, 1465.
(90) Dominy, B. N.; Brooks, C. L., III. *J. Phys. Chem.* **1999**, *103*, 3765.
(91) Onufriev, A.; Bashford, D.; Case, D. A. *J. Phys. Chem.* **2000**, *104*, 3712.
(92) Lee, M. S.; Salsbury, F. R., Jr.; Brooks, C. L., III. *J. Chem. Phys. B* **2002**, *116*, 10606.
(93) Lee, M. S.; Feig, M.; Salsbury, F. R., Jr.; Brooks, C. L., III. *J. Comput. Chem.* **2003**, *24*, 1348.
(94) Zhang, W.; Hou, T.; Qiao, X.; Xu, X. *J. Phys. Chem. B* **2003**, *107*, 9071.
(95) Jayaram, B.; Sprous, D.; Liu, Y.; Beveridge, D. L. *J. Chem. Phys. B* **1998**, *102*, 9571.
(96) Morozov, A. V.; Kortemme, T.; Baker, D. *J. Phys. Chem. B* **2003**, *107*, 2075.
(97) Li, Z.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 6611.
(98) von Freyberg, B.; Braun, W. *J. Comput. Chem.* **1991**, *12*, 1065.
(99) Liu, D. C.; Nocedal, J. *Technical Report NAM03*; Department of Electrical Engineering and Computer Science, Northwestern University: Evanston, IL, 1988.
(100) Flory, P. J. *Statistical Mechanics of Chain Molecules*; Hanser Publishers: New York, 1989.
(101) Muller, C. W.; Schulz, G. E. *J. Mol. Biol.* **1992**, *224*, 159.
(102) Hassan, S. A.; Guarnieri, F.; Mehler, E. L. *J. Phys. Chem. B* **2000**, *104*, 6478.
(103) Hassan, S. A.; Mehler, E. L. *Proteins* **2002**, *47*, 45.
(104) Warshel, A.; Russel, S. T. Q. *Rev. Biophys.* **1984**, *17*, 283.
(105) Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L. *J. Phys. Chem. B* **2001**, *105*, 6476.
(106) Blanco, F. J.; Rivas, G.; Sranno, L. *Nat. Struct. Biol.* **1994**, *1*, 584.
(107) Munoz, V.; Thompson, P. A.; Hofrichter, J.; Eaton, W. A. *Nature* **1997**, *390*, 196.
(108) Munoz, V.; Henry, E. R.; Hofrichter, J.; Eaton, W. A. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 5872.
(109) Pande, V. S.; Rokhsar, D. S. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 9062.
(110) Garcia, A. E.; Sanbonmatsu, K. Y. *Proteins* **2001**, *42*, 345.
(111) Zhou, R.; Berne, B. J.; Germain, R. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 14931.

(112) Zhou, R. *Proteins* **2003**, *53*, 148.
(113) Zhou, R.; Berne, B. J. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12777.
(114) Felts, A. K.; Harano, Y.; Gallicchio, E.; Levy, R. M. *Proteins* **2004** *56*, 310.
(115) Fogolari, F.; Tosatto, S. C. E. *Protein Science* **2005**, *14*, 889.
(116) Rapp, S. R.; Pollack, R. M. *Proteins* **2005**, *60*, 103.
(117) Kollman, P.; Dixon, R.; Cornell, W.; Fox, T.; Chipot, C.; Pohorille, A. The development/application of a "minimalist" organic/biochemical molecular mechanic force field using a combination of ab initio calculations and experimental data. In *Computer Simulation of Biomolecular Systems*; van Gunsteren, W. F., Weiner, P. K., Wilkinson, A. J., Eds.; Kluwer: Boston, MA, 1997; Vol. 3, p 83.
(118) Lee, M. S.; Salsbury, F. R., Jr.; Olson, M. A. *J. Comput. Chem.* **2004**, *25*, 1967.
(119) Lee, M. S.; Olson, M. A. *J. Phys. Chem. B* **2005**, *109*, 5223.
(120) Fan, H.; Mark, A. E.; Zhu, J.; Honig, B. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 6760.