Solvation Parameters for Predicting the Structure of Surface Loops in Proteins: Transferability and Entropic Effects

Bedamati Das and Hagai Meirovitch*
Center for Computational Biology and Bioinformatics and Department of Molecular Genetics and Biochemistry, School of Medicine, University of Pittsburgh, Pittsburgh, Pennsylvania

ABSTRACT **A new procedure for optimizing parameters of implicit solvation models introduced by us has been applied successfully first to cyclic peptides and more recently to three surface loops of ribonuclease A (Das and Meirovitch, Proteins 2001; 43:303–314) using the simplified model $E_{tot} = E_{FF}(\varepsilon = nr) + \Sigma_i \sigma_i A_i$, where σ_i are atomic solvation parameters (ASPs) to be optimized, A_i is the solvent accessible surface area of atom i, $E_{FF}(\varepsilon = nr)$ is the AMBER force-field energy of the loop-loop and loop-template interactions with a distance-dependent dielectric constant, $\varepsilon = nr$, where n is a parameter. The loop is free to move while the protein template is held fixed in its X-ray structure; an extensive conformational search for energy minimized loop structures is carried out with our local torsional deformation method. The optimal ASPs and n are those for which the structure with the lowest minimized energy $[E_{tot}(n,\sigma_i)]$ becomes the experimental X-ray structure, or less strictly, the energy gap between these structures is within 2–3 kcal/mol. To check if a set of ASPs can be defined, which is transferable to a large number of loops, we optimize individual sets of ASPs (based on $n = 2$) for 12 surface loops from which an "averaged" best-fit set is defined. This set is then applied to the 12 loops and an independent "test" group of 8 loops leading in most cases to very small RMSD values; thus, this set can be useful for structure prediction of loops in homology modeling. For three loops we also calculate the free energy gaps to find that they are only slightly smaller than their energy counterparts, indicating that only larger n will enable reducing too large gaps. Because of its simplicity, this model allowed carrying out an extensive application of our methodology, providing thereby a large number of benchmark results for comparison with future calculations based on $n > 2$ as well as on more sophisticated solvation models with as yet unknown performance for loops. Proteins 2003;51:470–483.**

INTRODUCTION

Surface loops in proteins take part in protein-protein and protein-ligand interactions, where their flexibility in many cases is essential for these recognition processes. The loop flexibility is demonstrated in multidimensional nuclear magnetic resonance (NMR) and X-ray crystallography experiments, where for the latter it is reflected in terms of large B-factors[1] or a complete disorder. Thus, the conformational change between a free and a bound antibody shows 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 (see, e.g., Refs. 2 and 3). Alternatively, the selected-fit mechanism has been suggested, where the free active site interconverts among different states, and one of them is selected upon binding[4]; the same also applies for loops. Dynamic NMR experiments 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 review by Fetrow[8]).

The interest in surface loops has yielded extensive theoretical work where one avenue of research has been the classification of loop structures.[8–16] However, to understand various recognition mechanisms like those mentioned above, it is mandatory to be able to predict the structure (or structures) of a loop by theoretical/computational procedures. As discussed in detail below, this is not a trivial task due to the irregular structures of loops, their flexibility, and exposure to the solvent, which requires developing adequate modeling of solvation. In fact, structure prediction of loops constitutes a challenge in protein engineering, where a loop undergoes mutations, insertions, or deletions of amino acids. Determination of the structure of large loops is still an unsolved problem in homology modeling.[17–19]

Loop structures are commonly predicted by a comparative modeling approach based on known loop structures

Grant sponsor: National Institutes of Health; Grant number: ROIGM61916.

*Correspondence to: Hagai Meirovitch, Center for Computational Biology and Bioinformatics and Department of Molecular Genetics and Biochemistry, School of Medicine, University of Pittsburgh, PA 15213. E-mail: hagaim@pitt.edu

Received 12 August 2002; Accepted 21 November 2002

from the Protein Data Bank (PDB),[20,21] an energetic approach, or methods that are hybrid of these two approaches. The first approach is based in most cases on matching segments from the database with the length of a target loop and the relative positions of its adjacent residues. Hence, this approach is especially appropriate for homology studies, where the protein framework is not known exactly. However, only short loops (up to five residues) could be treated effectively by comparative modeling.[22–26] To date, statistical and hybrid methods cannot handle loops of more than $n = 9$ residues because of the lack of sufficiently large databases[27–30] (see a detailed discussion in our previous article, Ref. 31).

With the energetic 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 three-dimensional (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.[27–39] These methods are not expected to handle efficiently large loops because of the lack of conformational search capabilities, which in many cases is partially caused by complex construction procedures based on at least two stages, where the side-chains are added to an initially generated backbone (see a more detailed discussion in our Ref. 31).

Modeling of the solvent is of special importance. In some of these studies, the solvation problem is not addressed at all, whereas most of them only use a distance-dependent dielectric constant ($\varepsilon = r$). Better treatments of solvation were applied by Moult and James[28] and Mas et al.[40] A systematic comparison of solvation models was first carried out by Smith and Honig[41] who tested the $\varepsilon = r$ model against results obtained by the Finite Difference Poisson Boltzmann calculation including a hydrophobic term; the implicit solvation model of Wesson and Eisenberg[42] with $\varepsilon = r$ was also studied by them. More recently, the GB/SA model[43] was applied to loops of Ribonuclease (RNase) A.[44] Comparing the efficiency of these methods is not straightforward. However, as expected, the structure prediction is found to improve as the ratio (loop length)/(distance between ends) decreases, and the conformational restrictions imposed by the template increase. For example, Bruccoleri et al.[45] obtained RMSD values within the range 0.7–2.6 Å for backbone atoms and 1.4–4.1 Å for all atoms for the 12 hypervariable loops (of size, 5–12 residues) of the antibodies McPC603 and HyHEL-5; these values are typical for such less restricted systems.[33] Although this imperfection has been attributed to the inadequate modeling of solvation, the better treatments of Smith and Honig[41] have also been found to be inconclusive. The results of Rapp and Friesner[44] have shown strong dependence on the force field. This supports our point of view that solvation parameters should be optimized together with the specific force field used. Thus far, none of these approaches has addressed the problem of loop flexibility in a systematic way.

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, and it constitutes the only alternative for studying the flexibility of loops. Recently, we developed a statistical mechanics methodology[46–49] for treating flexibility, which was used successfully to predict the solution structures and populations of cyclic peptides in DMSO.[50,51] This methodology relies on a novel and general method for optimizing parameters of implicit solvation models, which was used to optimize atomic solvation parameters (ASPs), σ_i, of the simplified well-studied solvation model,

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

where E_{FF} is the force-field energy and A_i is the solvent-accessible surface area of atom i; $\varepsilon = nr$ is a distance-dependent dielectric constant where n is an additional parameter. This optimization requires extensive conformational search for low-energy minimized structures, which is carried out with our highly efficient local torsional deformations (LTD) method.[46,52,53] From a large sample of structures generated by LTD (using E_{tot} with the optimized ASPs) one identifies a relatively small group of low-energy structures that are significantly different. Each of the latter becomes a "seed" for Monte Carlo simulation which spans its vicinity, the free energies (hence the relative populations) are calculated from the MC samples with the local states (LS) method,[54,55] and averages of various properties over the samples' contributions weighted by the populations can be calculated and compared with the experiment. Our long-range objective is to extend this methodology to surface loops of proteins in water, and an initial step in this direction was carried out in our previous article[31] (called here article I), where the optimization procedure [based on Eq. 1 with the all-atom AMBER force field (E_{FF})] was tested as applied to two surface loops of RNase A, leading to optimal $n = 2$ and two independent but similar sets of ASPs.

It should be pointed out that although our parameter optimization is general, its practical applicability depends on the complexity of the solvation model and the loop size. Therefore, we have initially chosen to treat the relatively simplified model defined in Eq. 1 (see article I), which would allow studying various technical problems as well as the quality of the solvation model; moreover, the calculations are expected to provide a set of benchmark results for comparison with results based on more sophisticated models. However, even for this simplified model, application of the methodology is feasible only for a relatively small template, including those atoms that are located within 10 Å from any loop atom; these atoms are fixed in their known X-ray structure, whereas the loop is free to move. E_{tot} includes the loop-loop and loop-template energy, whereas the template-template interactions are ignored. The optimized ASPs and n are those for which the known X-ray loop structure becomes the global energy minimized (GEM) structure. This definition, however, turns out to be too strict and in article I 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 after local energy minimization of the X-ray loop structure; see Results and Discussion).

Using the molecular mechanics/molecular dynamics program TINKER[56] we have also checked in article I the all-atom OPLS[57] and CHARMM22[58] force fields and have found that for loops the AMBER[59] force field performs the best; hence, the ASPs were optimized with respect to AMBER. For the optimized sets of ASPs (denoted σ_i^*) and the optimal $n = 2$, energy gaps, $\Delta E_{\text{tot}}^m(n, \sigma_i^*)$ of 1.9 and 1.1 kcal/mol were found for loop 1 (64–71) and 3 (89–97), respectively, where $\Delta E_{\text{tot}}^m(n,\sigma_i^*)$ is

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

and $E^m{}_{\text{tot}}(n,\sigma_i^*)$ is the lowest minimized energy obtained, which is assumed to be the GEM. The GEM loop structures were found to have relatively small root-mean-square deviations (RMSD) with respect to the corresponding NOS structures, not only for the backbone but also for the side-chains. An important factor in these calculations has been the efficiency of LTD, which unlike other conformational search methods for loops handles the whole loop (i.e., backbone and side-chains) at once, whereas the loop closure is carried out automatically by the energy minimization. It should be pointed out that we are interested in the solution behavior of loops, whereas the optimization is carried out with respect to the crystal structure; therefore, in this analysis the flexibility of the loops as reflected in NMR[60] and X-ray experiments[61–63] was also taken into account. The fact that these independent sets of ASPs are similar has enabled defining a meaningful averaged best-fit set of ASPs that led to small RMSD values for both loops as well. In summary, this derivation of ASPs depends on the force field used and is based on the energy of the entire loop in the protein environment, in contrast to the conventional parametrization that relies on free energy of transfer data of small molecules.[42]

However, despite the very encouraging results of article I, one should still verify that the parameter optimization can be applied to other loops of different size and sequence of amino acid residues from various proteins and that a transferable best-fit set of ASPs can be defined. This study is carried out in the present article where independent sets of ASPs are optimized (based on the AMBER force field and using TINKER) for a diverse group of 12 surface loops of size 5–12 residues. The extent of similarity among these individual sets enables defining a reasonable best-fit set of ASPs, which is tested as applied to the original group of 12 loops as well as to an additional group of different eight loops. Because the free energy, rather than the minimized energy, constitutes the correct criterion of stability, we study for the first time the free energy differences between the wide microstates (see next section) of the GEM and NOS structures for three loops. Thus, MC simulations are started from the corresponding energy-minimized structures, and the free energy is calculated with the LS method from the generated samples (wide microstates).

THEORY AND METHODS

In this section we describe the methodology for treating flexibility, the treatment of solvent effects, and the LS method for calculating the free energy.

Methodology for Treating Flexibility

A long non-stretched surface loop surrounded by a constant protein template typically spans a large number of energy-minimized structures, where the ensemble of conformations pertaining to the basin of attraction of each minimum (i.e., those that would reach this minimum by local energy minimization) is called a localized microstate. The energy landscape of the loop also contains larger potential wells defined over regions called wide microstates, where each consists of many localized ones (for a protein a wide microstate is defined, e.g., by the local MD fluctuations around its averaged structure; for an operative definition of a wide microstate by MC or MD simulations, see Ref. 49). MD studies have shown that the molecule will visit a localized microstate only for a very short time (several femtoseconds) while staying for a much longer time within a wide microstate,[64–66] which means that the wide microstates are of a greater physical significance than the localized ones. Thus, the experimental data in solution obtained for a loop with a well-defined structure should be compared with theoretical values averaged over the most stable wide microstate, which is defined by the local loop fluctuations simulated by MC or MD. A large surface loop might also be a random coil or exhibit an intermediate flexibility between these two extreme cases, where several wide microstates are populated significantly in thermodynamic equilibrium.

To determine the extent of flexibility one should identify the most stable wide microstates i, i.e., those with the largest contribution Z_i to the total partition function of the loop. The relative populations, $p_i = Z_i/\Sigma Z_i$ then lead to the statistical average of a property G, $\langle G \rangle = p_i G_i$, where G_i is the contribution of wide microstates i. Identification of the most stable wide microstates is achieved in two stages (see Refs. 46–49 for peptides and article I for loops). First, by using LTD, an extensive conformational search is carried out for the global energy minimized (GEM) structure and low-energy minimized structures within 2–3 kcal/mol above the GEM. These minimized structures should reside within the most stable wide microstates, and a subgroup of them that are significantly different would represent the different wide microstates, because per definition, structures that pertain to the same wide microstate are similar (see discussions in article I).

A suitable criterion of variance for two structures is that at least one dihedral angle differs by 60° or more. This angular criterion, which is based on energetic considerations, has been found to be suitable for a short peptide, whereas for a long peptide or loop, an additional criterion, such as the RMSD between structures should be employed (see discussion in Ref. 53). Each selected structure then becomes a "seed" for an MC or MD simulation that spans the related wide microstate. The free energies, F_i, of the most stable wide microstates are obtained with the LS

method applied to the corresponding samples. Criteria developed previously[49,47] enable one to check the structural distinctiveness and thermodynamic stability of the various samples, (i.e., that they do not overlap and remain in their original conformational regions). As mentioned in the Introduction, in article I we have optimized sets of ASPs for two loops of RNase A, but without calculating the free energy of the most stable wide microstates; a preliminary study is carried out here.

Local States Method

The LS method has been discussed in detail in previous publications as applied to peptides and loops[47,49,54,55,67]; therefore, we shall give only a brief description here. This method enables one to calculate approximately the free energy of a macromolecule from a given sample of conformations generated by any simulation technique, in particular the MC and MD methods. The K dihedral and bond angles of the backbone and side-chains are denoted by α_k, where k defines their order along the chain. In the first step, one calculates for each wide microstate the variability range $\Delta\alpha_k$

$$\Delta\alpha_k = \alpha_k(\max) - \alpha_k(\min) \tag{3}$$

where $\alpha_k(\max)$ and $\alpha_k(\min)$ are the maximum and minimum values of α found in the sample, respectively. Next, the ranges $\Delta\alpha_k$ are divided into l equal segments, where l is the discretization parameter. We denote these segments by v_k, ($v_k = 1,l$). Thus, an angle α_k is now represented by the segment v_k to which it belongs and a conformation i is expressed by the corresponding vector of segments $[v_1(i), v_2(i), \ldots, v_k(i)]$. A local state related to v_k is a partial conformation which consists of v_k and the b angles preceding it along the chain, that is, the vector $(v_k, v_{k-1}, \ldots, v_{k-b})$; b is called the correlation parameter. For a given b, one calculates from the sample the number of occurrences $n(v_k, v_{k-1}, \ldots, v_{k-b})$ of all the local states from which transition probabilities $p(v_k|v_{k-1}, \ldots, v_{k-b})$ can be defined. Now, for each member i of the sample one determines the K local states and the corresponding transition probabilities, whose product defines an approximate probability density $\rho_i(b,l)$ for conformation i

$$\rho_i(b,l) = \prod_{k=1}^{K} p(v_k|v_{k-1}, \ldots, v_{k-b})/(\Delta\alpha_k/l). \tag{4}$$

$\rho_i(b,l)$ allows defining an approximate free energy functional F^A which constitutes a rigorous lower bound for the correct free energy, F,

$$F^A(b,l) = \langle E\rangle + k_B T \sum_i \rho_i(b,l) \ln \rho_i(b,l) \tag{5}$$

where $\langle E\rangle$ is the Boltzmann average of the potential energy, estimated from the MD or MC sample. This enables one to calculate a lower bound for the free energy, $F^A_m(b,l)$ for each wide microstate m. The true difference in the free energy $\Delta F_{m,n}$ between wide microstates m and n is obtained from the convergence of the differences

$$\Delta F^A_{m,n}(b,l) = F^A_m(b,l) - F^A_n(b,l) \tag{6}$$

for many improving approximations, (i.e., larger values of b and l). The LS method can be applied to any chain flexibility (i.e., it is not limited to harmonic or quasi-harmonic fluctuations).[65] Thus, free energy differences between wide microstates with significant structural differences can be calculated, which is a difficult task with methods based on thermodynamic integration.

Modeling Solvation Effects

An essential ingredient of this methodology is a reliable energy function, which takes into account solvent effects. Because explicit water, the most accurate model, is computationally expensive, we have chosen to study the relatively simple implicit solvation model defined in Eq. 1. With this model, the conformational search, the identification of the most stable wide microstates, and the calculation of their free energy is considerably easier than with explicit solvent; in particular, calculating the difference in the free energy of significantly different wide microstates with the commonly used perturbation and thermodynamic integration techniques[68,67] is relatively inefficient for explicit water. Therefore, most of the loop studies in the literature are based on implicit solvation models and the work of McCammon's group[69] on loops of the anti-insulin antibody using explicit water is an exception (see also Ref. 25),

Eq. 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.[42,70] However, it is not clear to what extent ASPs derived for small molecules are suited for the protein environment. In addition, these sets of ASPs were used with various force fields, in most cases without further calibration (see discussions in Refs. 46 and 49 and in references cited therein). This seems unjustified because the existing force fields are different, and probably none of them is expected to faithfully describe a protein in vacuum; therefore, even if a set of ASPs has been derived, which describe correctly the first hydration shell of a protein, E_{tot} would still be inaccurate. In other words, the ASPs should be optimized with respect to the force-field energy used. Recent studies based on various solvation potentials, E_{solv}, support these reservations,[41,44] including our results in article I. This problem was first noticed by Schiffer et al.,[71] and more recently by Fraternali and van Gunsteren.[72] It should be emphasized again that our optimization procedure is applicable to any solvation model, and the choice of the present model (Eq. 1) stems from its simplicity and popularity. Thus, the present calculations will also provide benchmark results for comparison with future calculations based on more sophisticated solvation models (see Summary).

One would expect the optimal ASPs to express to a certain extent the Born self–energies[73] and the hydrophobic interactions. The screening of the electrostatic interactions by the surrounding water is modeled approximately by a distance-dependent dielectric constant, $\varepsilon = nr$. Notice

TABLE I. Proteins, Loops, and Their Templates[†]

Protein	Loop	Sequence	R	No. of atoms (loop)	No. of atoms (temp.)	Radius (Å) (temp.)
Antibody, McPC603 (1mcp)	Loop 1, L26–L37 (12)	SQSLLNSGNQKN	2.5	175	716	9
Antibody, McPC603 (1mcp)	Loop 2, H102–H109 (8)	YYGSTWYF	3.7	139	975	7
RNase A (1rat)	Loop 3, 89–97 (9)	SSKYPNCAY	2.8	133	593	10
Proteinase (2apr)	202–210 (9)	ATVGTSTVA	4.8	112	692	9
Acidic fibroblast (FGF) (2afg)	90–94 (5)	EENHY	2.3	84	552	10
Adenylate kinase (AK) (4ake)	73–80 (8)	AQEDCRNG	2.1	112	605	9
Penicillopepsin (3app)	129–137 (9)	INTVQPQSQ	2.7	139	863	9
Peptidase (5cpa)	205–213 (9)	PYGYTTQSI	3.5	138	899	9
Myoglobin (1mba)	119–125 (7)	SVAAPPA	1.1	85	710	10
BPTI (8pti)	Loop 1, 6–12 (7)	LEPPYTG	~1	105	682	10
BPTI (8pti)	Loop 2, 18–24 (7)	IIRYFYN	~1	138	740	10

[†]R is the ratio between the length of the stretched loop and the distance between the C$^{\alpha}$ of the first and last residues of the loop.

that E_{tot} is a free energy function that depends on the temperature (through the σ_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 because of strong hydrophilic interactions. Indeed, the ASPs optimized for the two loops of RNase A in article I are all negative (hydrophilic), even those of carbon (in contrast to the positive ASP obtained by, e.g., Wesson and Eisenberg[42]). As explained there, the hydrophobic interactions have been already taken into account in creating the protein's native structure, and thus the negative ASP of carbon is probably due to its self-energy $\sim -q^2/R_c$, where q is the partial charge of carbon and R_c is its Born radius. Therefore, the effect of the loop's carbons (which constitute most of the atoms) is to "help" the hydrophilic atoms define (against the tendency of E_{FF}) lower E_{tot} for loop structures that protrude into the solvent than for those that collapse on the template. This means that the ASPs would not be effective for highly stretched loops that cannot change their conformation in response to the attracting solvent, as demonstrated for the 12-residue loop 2 (13–24) of RNase A studied in article I. The ASPs are expected to become ineffective also for loops that are prevented from moving outward because of other geometrical constraints imposed by the protein's template, as discussed later.

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.[74,75] On the other hand, 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 situation is currently being changed, however, where high-resolution X-ray structures become available, which enables extracting information about side-chain rotamers and their populations.[76,77] For example, in the high-resolution crystal structure (0.87 Å at 100 K) of RNase A, where Asn^{67} was replaced by an isoaspartyl residue, 15% of the side-chains (19) were found to populate multiple rotamers. However, such information is not available for most of the loops studied here. In addition, the derivation of the ASPs is based on the minimized energies, thus ignoring the local flexibility of the corresponding wide microstates. The first step to correct this limitation is done here, where differences in the free energy for three loops are calculated.

RESULTS AND DISCUSSION

In this section, based on Eq. 1, sets of ASPs are optimized independently for 12 surface loops, a best-fit set is derived, and the quality of these sets is discussed. We also carry out several MC simulations of GEM and NOS wide microstates and calculate the differences in their free energy using the LS method. Finally, the best-fit set is tested as applied to an independent group of eight loops.

Loops Studied and Modeling Issues

The 12 surface loops for which individual sets of ASPs are optimized and the related proteins are presented in Table I. The 3D structures of these proteins, taken from the Protein Data Bank (PDB), all have been determined with 2 Å resolution or less, beside that of the antibody McPC603 that was obtained with 2.7 Å resolution. To be able to compare our results with those obtained by other methods, most of the loops selected were studied previously besides the loops of adenylate kinase (AK) and acidic fibroblast (FGF). The 12 loops range in size from 5 to 12 amino acid residues, where most of them are predominantly hydrophilic (i.e., polar or charged); an exception is the loop of myoglobin, which is composed of 6 (of 7) hydrophobic (non-polar) residues. 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) have been obtained with elevated B-factors. In particular, for AK, the B-factors of CD, OE1, and NE2 of Gln^{74} are 57–64, the B-factors of all the atoms of Glu^{75} range from 48 to 88, and those for the side-chain atoms of Arg^{78} from 47 to 85. For chain B of acidic FGF (see below), OE1 of Glu^{90} has a B-factor of 75, the B factors of CG, CD, and OE1 of Glu^{91} are 50, 100, and 100, respectively, and that of O(H) of Tyr^{94} is 65. These large B-factors suggest that the side-chains might populate several rotamers, but in contrast to the case of RNase A, no analysis of rotamers' population is available [For example, Müller and Schulz do not determine dihedral angles if the B-

factors of the involved atoms are 60 and above,[78] whereas others adopt 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 ASPs.

The ASPs 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. To check whether the loops of molecule A of AK and molecule B (FGF) studied by us are influenced by the electrostatic interactions with the neighbor molecules, we calculated the minimal distance between all the atoms of a loop and the atoms of all the neighbor molecules. For the AK loop, this minimal distance is relatively large, 14.6 Å, whereas it is only 3.5 Å for the acidic FGF's loop. Therefore, for acidic FGF we have also studied the loop 90–94 of molecule C, which has a relatively large minimal distance of 22.5 Å and lower B-factors than those of the loop of molecule B. Thus, the largest B-factors for molecule C are 67 for CG and OE1 of Glu[91] and 91 for ND2 of Asn.[92] Correspondingly, as discussed later, the energy gaps for these two acidic FGF loops are different. The ASPs might also be affected from close molecules in neighbor cells; however, we have not attempted to check this point.

Another parameter presented in Table I is the ratio R = [length of the stretched loop/distance between its ends], which is calculated between the C^{α} atoms of the first and last residues of the loop. The length (in Å) of the extended structure is calculated by 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's book[79] (Chapter VII, p. 251). To a large extent, R reflects the conformational freedom of the loop's backbone and to a lesser extent also that of the side-chains, the larger R is the higher the flexibility; for convenience, these R-values are also provided in Table III. The table reveals that the loops of myoglobin and BPTI are strongly stretched, which explains the high rate of success achieved by various methods in predicting their structure (see the Results section for the RMSD).

The table provides the number of atoms (including hydrogens) of each loop, which ranges from 84 (acidic FGF) to 175 (the 12-residue loop of the antibody). We also present in the table the number of atoms in each template, which is defined by the following procedure: First, hydrogen atoms are added to the PDB X-ray structure by the program TINKER. To remove possible atomic overlaps, the energy of the protein is minimized by using the AMBER potential [$E_{FF}(\varepsilon = 1)$] with an additional harmonic restraint of 5 kcal/mol/Å^2 applied to each atomic position. This minimized structure is called the native optimized structure (NOS), which deviates from the PDB structure by an all-heavy-atom RMSD of ~0.15 Å. The templates defined in article I include any non-loop atom with a distance < 10 Å from at least one loop atom (in NOS) together with all the other atoms pertaining to the same residue. However, some of the proteins studied here are significantly larger than RNase A, and using a distance of 10 Å would increase significantly the template and the required computer time for the energy calculations. Therefore, for these proteins we used a distance (denoted radius in Table I) < 10 Å. However, we have found that decreasing the distance from 10 to 7 changed the energy only slightly (≤1 kcal/mol), suggesting that the effect on differences in energy between two structures would be small. The template sizes range from 552 (acidic FGF) to 975 (antibody, loop 2).

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 article I it was found necessary to optimize the orientations of the OH and NH vectors of NOS and the template. This was carried out by a Monte Carlo minimization procedure, where the polar vectors were 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 (for details see Appendix C of article I). As discussed in article I, this optimization is more important for the single polar hydrogens (e.g., the OH group of the side-chain of Ser) becoming less important for Lys (e.g., where three NH groups are rotated together). These optimizations of the polar hydrogen networks [using $E_{FF}(\varepsilon = nr)$] lead to NOS structures that deviate by RMSD ~0.2 Å from the PDB loop structures; in this work (as in article I) these NOS structures are considered to be the correct (experimental) ones against which the RMSD of structures is calculated.

As in article I, the energy was minimized by the L-BFGS procedure,[80] the solvent-accessible surface area (SASA) and its first derivatives were calculated by the program MSEED[81] for both the loop and the template, and the conformational search was carried out with the LTD method,[52,31] where all these programs have been incorporated within TINKER. For calculating the SASA, a water molecule is represented by a sphere of radius 1.4 Å, and the radius r_i of atom i, is determined from its Lennard-Jones parameter σ_i(LJ), where $r_i = 2^{1/6}\sigma_i(\mathrm{LJ})/2$; the radius of a hydrogen is 0.9 Å. Only for the calculation of SASA, CH, CH2, and CH3 are treated as united atoms with the same radius of 2.1 Å.

In article I we used the all-atom AMBER force field where the five residues, Arg, Lys, His, Asp, and Glu, are charged, obtaining very good results for loops 1 and 3, which both contain a single Lys residue. However, applying this force field to the loops of Table I, especially to the highly charged loops of AK and acidic FGF has led to large energy gaps (>20 kcal/mol), which could not be decreased sufficiently even with relatively large distance-dependent dielectric constants $\varepsilon = nr$, where n = 3 to 6. Therefore, we have decided to use electrostatically neutralized models of these residues where four of them are provided by AMBER, whereas Arg has been neutralized by us using partial charges of −0.6161 for NE, +0.4 for CZ, −0.7 for NH1 and NH2, and +0.24 for each of the four hydrogens connected to NH1 and NH2; the rest of the charges of the residue have not been changed; we have also neutralized the charges of the end groups of the protein. Although the neutralized residues have led to a significant reduction in

TABLE II. ASPs and the Energy Gaps[†]

	ASPs [cal/(mol · Å^2)]						Gap (kcal/mol)		
Protein/loop	1 ASP	C	N	O	H	S	FF	1 ASP	ASPs
Antibody	−0.08	−0.04	−0.10	−0.10	−0.08	−0.08	14.6	8.2	4.9
Loop 1, L26–37 (12)		−0.04	−0.10	−0.10	−0.09	−0.08	14.6		4.8
Antibody	−0.09	−0.03	−0.08	−0.09	−0.12	−0.12	5.5	2.3	1.6
Loop 2, H102–109 (8)							5.5		1.9
RNase A	−0.05	−0.04	−0.09	−0.05	−0.05	−0.05	5.5	2.0	1.8
Loop 3, 89–97 (9)							5.5		1.9
Proteinase	−0.06	−0.04	−0.12	−0.20	−0.12	−0.06	4.9	3.6	0.5
202–210 (9)							4.9		3.4
Acidic FGF	−0.09	−0.09	−0.09	−0.09	−0.09	−0.09	12.3	4.6	4.6
90–94 (5)							12.3		8.5
AK	−0.12	−0.06	−0.12	−0.06	−0.12	−0.12	11.8	9.1	6.0
73–80 (8)							11.8		13.5
Penicillopepsin	−0.14	+0.01	−0.18	−0.15	−0.13	−0.13	10.3	4.8	1.8
129–137 (9)							10.3		4.1
Peptidase	−0.02	+0.06	+0.11	−0.06	−0.05	−0.02	4.1	4.1	0.5
205–213 (9)							4.1		6.0
Myoglobin	+0.04						1.0	0.6	
118–124 (7)									6.1
BPTI							0.4		
Loop 1, 6–12 (7)							0.4		0.1
BPTI							0.0		
Loop 2, 18–24 (7)							0.0		0.3

[†]In the first line of each loop appear the optimized ASPs and the corresponding energy gaps; in the second line of loop 1 of the antibody appear the best-fit set of ASPs and in the second line of each loop the energy gaps based on the best-fit set. The ASP for S was not optimized and the value used is provided. FF, 1 ASP, and ASPs are the energy gap obtained with the force field alone, one ASP and the set of ASPs, respectively.

the energy gaps, it should be pointed out that their electrostatic interactions are still significant due to large dipole moments. This system neutralization, besides its practical necessity, also has theoretical justification: protein systems have neutral charge due to counterions.

The optimization of the ASPs is based on a conformational search for low-energy minimized structures carried out with LTD in several stages, as described in detail in article I. Thus, using the force field [$E_{FF}(\varepsilon = nr)$] alone, the GEM structure of the loop is found and the energy gap ΔE^m_{tot} $(n,\sigma_i = 0)$ (Eq. 2) with respect to the minimized energy of NOS is calculated (notice, however, that GEM denotes here the lowest energy obtained, which still might not be the global minimum); this procedure consists typically of 3000 energy minimizations. In the next stage using E_{tot} (Eq. 1) a single ASP for all the atoms is optimized and finally a set of ASPs is determined for the various atoms. The entire process requires approximately 20,000 energy minimizations using TINKER. It should be pointed out that at each stage the optimization is based on ~600 significantly different structures (according to the 60° criterion) with minimized energies within a ~7 kcal/mol range above the GEM; this group is changed in the course of the optimization (see Appendix B of article I).

Because of the neutralization of the charges discussed above we had to check whether the choice $\varepsilon = 2r$ used in article I is still valid. Therefore, we carried out optimizations of ASPs with n = 1.5, 2, and 3 (and for AK also for larger n, up to 6). It should be pointed out that the difference in the force-field energy, E_{FF}, between two structures is due to a large extent to the electrostatic interactions, which are highly sensitive to conformational changes; therefore, if these interactions are decreased by neutralizing the charges or increasing the dielectric constant, E_{FF} becomes less sensitive to conformational changes. Indeed, for $n = 3$, where the electrostatic interactions are relatively muted, we could not optimize ASPs for several loops, (i.e., the same energy gap has been obtained for a wide range of ASPs' values). Therefore, we have used in this work (as in article I), $n = 2$, which leads to a stronger decrease of the electrostatic interaction than n = 1.5 but still enables optimizing the ASPs.

Results for the ASPs

The optimized ASPs and the corresponding energy gaps for E_{FF} and E_{tot} [Δ^m_{tot} (n,σ^*_i), Eq. 2] appear in Table II. It should first be pointed out that for the first seven loops the single ASP is negative, leading to significantly smaller energy gaps than those obtained by the force field alone, meaning that the NOS structures protrude into the solvent. The sets of ASPs are negative as well including the ASPs of carbon (except for penicillopepsin) as was found in article I, and the energy gaps are smaller than the corresponding gaps based on a single ASP. The four last loops require special discussion. Loop 2 of BPTI is highly stretched (see Table I), and NOS has been found to be the GEM structure already for the force-field energy; therefore, no attempt has been made to optimize ASPs for this loop. Loop 1 of BPTI is also stretched but somewhat less than loop 2, leading to a relatively small energy gap of 0.4 kcal/mol for E_{FF}; therefore, ASPs were not optimized for this loop as well. The loop of myoglobin is again highly

TABLE III. Heavy Atoms RMSD of the GEM Structures From the NOS Structures[†]

		RMSD-FF (Å)			RMSD-1 ASP (Å)			RMSD-ASPs (Å)		
Protein/loop	R	BB	SC	TOT	BB	SC	TOT	BB	SC	TOT
Antibody	2.5	1.1	2.3	1.8	1.4	2.6	2.0	1.0	1.7	1.4
Loop 1, L26–37 (12)								0.9	2.2	1.7
Antibody	3.7	1.2	0.9	1.0	1.1	0.9	0.9	0.7	0.7	0.7
Loop 2, H102–109 (8)								0.6	0.9	0.8
RNase A	2.8	0.5	1.5	1.1	0.4	1.5	1.1	0.3	1.2	0.9
Loop 3, 89–97 (9)								0.2	1.4	1.0
Proteinase	4.8	1.9	3.1	2.4	0.3	1.2	0.7	0.3	1.2	0.7
202–210 (9)								0.4	1.3	0.8
Acidic FGF	2.3	0.6	3.1	2.4	0.2	1.5	1.2	0.2	1.5	1.2
90–94 (5)								0.5	1.7	1.3
AK	2.1	1.1	3.5	2.5	0.7	2.2	1.6	1.0	3.2	2.3
73–80 (8)								0.9	2.9	2.1
Penicillopepsin	2.7	0.4	1.9	1.4	0.3	1.8	1.3	0.1	1.2	0.9
129–137 (9)								0.1	1.5	1.0
Peptidase	3.5	0.8	1.3	1.0	0.8	1.2	1.0	0.2	0.9	0.6
205–213 (9)								0.2	1.2	0.9
Myoglobin	1.1	0.2	0.7	0.5	0.3	0.8	0.5			
118–124 (7)								0.2	0.5	0.4
BPTI	~1.	0.1	1.5	1.0						
Loop 1, 6–12 (7)								0.1	0.8	0.5
BPTI	~1.	0.0	0.0	0.0						
Loop 2, 18–24 (7)								0.0	0.9	0.7

[†]R is defined in the caption of Table I. FF, 1 ASP, and ASPs denote results obtained with the force field alone, with a single ASP, and with a set of ASPs, respectively. BB, SC, and TOT are RMSD values for the backbone, side-chains, and their total, respectively. The results in the first and second line of each loop are for the optimized and best-fit set of ASPs, respectively.

stretched but less than loop 1 of BPTI and, therefore, the force field alone leads to a slightly larger gap of 1 kcal/mol. However, unlike the other loops, this loop is predominantly hydrophobic and it thus collapses on the template while the GEM [$E_{FF}(\varepsilon = 2r)$] structure is more open. Therefore, a lower gap of 0.6 kcal/mol was obtained with a single positive ASP of +0.4 cal/(mol · Å^2). Notice that further reduction in this gap energy could probably be achieved by larger ASP values, but we have not attempted to find the optimal value because our theory applies only to hydrophilic loops.

Another interesting case is the hydrophilic non-stretched ($R = 3.5$) loop of peptidase that thus seems to be a "normal" loop but unexpectedly has positive ASPs for C and N. However, we have checked by computer graphics the NOS and GEM [$E_{FF}(\varepsilon = 2r)$] structures and found that the first three residues, Pro, Tyr, and Gly, are located in a "pocket" created by the template, which does not allow them to move. The rest of the loop is relatively stretched with R = 1.3, and the conformations of Tyr^4, Thr^5, Thr^6, and Ile^9 in NOS and GEM overlap. The structural changes occur only for the backbone and side-chains of Gln^7 and Ser^8 (starting with a distortion of ψ of Thr^6), which leads to RMSD of 0.8 and 1.3 Å [based on $E_{FF}(\varepsilon = 2r)$] for the backbone and the side-chains, respectively (see Table III). Thus, the optimal ASPs rather than acting to decrease the relative energy of the entire NOS structure actually correct a particular localized arrangement of two residues only (resulting in RMSD 0.2 and 0.9 Å, respectively; see Table II), which happen to require positive ASPs for C and N. It should be pointed out that positive ASP of C [0.1 cal/(mol · Å^2] has been optimized also for the loop of penicillopepsin; however, the reason for this value is unclear.

As expected, the energy gaps decrease in going from the force field (FF), to one ASP, and to a set of ASPs; only for acidic FGF-B the same gap is obtained for one ASP and a set of ASPs. The gaps for the optimized sets of ASPs are all smaller than 2 kcal/mol besides those for the 12-residue loop of the antibody and the two highly charged loops of acidic FGF and AK with gaps of 4.9, 4.6, and 6 kcal/mol, respectively. Although these gaps are too large, one should bear in mind that the present procedure is based on the minimized energy, whereas the correct criterion for stability is the conformational free energy of the related wide microstates. Free energy differences are expected to be smaller than energy differences because in most cases, the wide microstate with the lower energy has lower entropy. In addition, as has been pointed out earlier, the X-ray coordinates of the loop of AK and to a lesser extent those of the loop of acidic FGF are not well defined (high B-factors), meaning that their NOS structures and the corresponding gaps are not well defined as well. Interaction of the loop's atoms with other molecules in the unit cell also can affect the results. Indeed, optimization of loop 90–94 of molecule C of acidic FGF, which is more distant from the other molecules in the unit cell than the loop of molecule B, has led to the lower energy gaps, 11.8 (vs 12.3), 4.1 (vs 4.6), and 3.8 (vs 4.6) kcal/mol for $E_{FF}(\varepsilon = 2r)$, a single ASP [−0.09 cal/(mol · Å^2)], and a set of optimized ASPs [$\sigma^*(C) = -0.10$, $\sigma^*(N) = \sigma^*(O) = \sigma^*(H) = -0.09$ cal/(mol · Å^2)], respectively (these results do not appear in Table II).

On the basis of the results for the sets of ASPs we have defined a best-fit set of ASPs that appears in Table II in the second line of loop 1 of the antibody. Because this set is similar to the optimized sets of the antibody's loops and the loop of RNase A, the best-fit gaps are only slightly larger than the optimized ones. As the difference between the optimized and best-fit ASPs increases, the best-fit gaps increase as well, as observed for proteinase and penicillopepsin, where the increase is from 0.7 to 3.4 and from 1.8 to 4.1 kcal/mol, respectively. Applying the best-fit set to loop 1 of BPTI decreased the gap from 0.4 (FF) to 0.1 kcal/mol, whereas the gap of loop 2 increased marginally from 0 to 0.3 kcal/mol because some distortion in the side-chains as is evident from Table III. As expected, the loops of peptidase and myoglobin are not treated adequately by the best-fit set of ASPs. The best-fit energy gaps for the loops of acidic FGF-B and AK are large, but the significance of these results is not clear due to the uncertainty in the correctness of the NOS structures, as has already been discussed.

Results for the RMSD

Results for the RMSD of the GEM structures from the corresponding NOS structures based on the heavy atoms and without superposition on the NOS structures are presented in Table III. As in Table II, the results in the second row in the columns denoted RMSD-ASPs have been obtained by the best-fit ASPs. A criterion for a successful prediction of the loop's backbone (BB) structure is that the RMSD from the correct structure is not larger than 1 Å[27,28]; notice, however, that RMSD values < 0.4 Å are actually insignificant because the two structures belong to the same wide microstate. For RMSD (FF), this criterion is not satisfied for four loops where the largest value of 1.9 Å is observed for the loop of proteinase, which correspondingly also has the largest R value. Notice, however, that although seven of the eleven loops satisfy this criterion, four of them (the loops of BPTI, myoglobin, and peptidase) are stretched and, therefore, do not provide support for the quality of the AMBER force field.

The RMSD (BB) values based on a single ASP are smaller in most cases than their counterparts obtained with the force field alone. The exception is loop 1 of the antibody, which probably stems from insufficient sampling for this relatively large loop. Notice that the ASP for myoglobin is not optimized and the increase of the RMSD from 0.2 to 0.3 Å is thus insignificant.

The results for RMSD (BB) based on the optimized set of ASPs are equal to or lower than those based on a single ASP; they are not > 1 Å, and in most cases are closer to 0 than to 1 Å. An exception is the loop of AK where the RMSD increases slightly from 0.7 to 1.0. It is interesting to point out that all of the backbone RMSD values obtained by the best-fit ASPs are < 1 Å and in some cases even smaller than or comparable to those obtained by the optimal sets of ASPs. As expected, treating the stretched loops of BPTI by the best-fit ASPs actually did not change their RMSD (FF) values.

The RMSD values of the side-chains are significantly larger than their backbone counterparts, ranging from 0.7 to 3.5 Å for the force-field calculations; for a single ASP, these values decrease (or remain unchanged) for seven loops where the largest RMSD reduces to 2.2 Å, and only for the antibody and myoglobin the RMSD values slightly increase. Further decrease (or unchanged) is observed in most cases for the RMSD values based on the optimal ASPs where excluding the result for loop 1 of the antibody (1.7 Å) and that of AK (3.2Å) the range reduces to 0.8–1.5 Å. The RMSD results based on the best-fit ASPs(bf), however, in most cases (besides for AK) are up to 0.5 Å larger than their counterparts obtained by the optimal ASPs. Although the highly stretched loop 1 of BPTI, as expected, has for the backbone a very small RMSD (FF) = 0.1 Å, for the side-chains RMSD(FF) = 1.5 Å, is relatively large; this value decreases to 0.8 Å using the best-fit set of ASPs. For the side-chains of loop 2 of BPTI, on the other hand, RMSD(bf) = 0.9 is slightly larger than RMSD (FF) = 0.0 Å. Notice that although RMSD(ASPs) for the side-chains are in most cases significantly better than those based on the force field, their values also reflect the uncertainty in the experimental structures (i.e., the high probability that a side-chain populates several rotamers). In article I, it is shown that if experimental information about side-chain flexibility is available, the RMSD can be decreased significantly. For the low-energy structures we have plotted the RMSD values against the corresponding minimized energies and obtained plots that are similar to those provided in article I; therefore, these plots are not shown here.

In this context, it should be pointed out that LTD, which has not been discussed in detail in this article, is a stochastic multistep method based at each step on local but significant rotations around dihedral angles followed by energy minimization, which enable the loop to cross energy barriers very efficiently. Thus, a significant part of conformational space is sampled as expressed by the large maximal RMSD (TOT) values obtained from the LTD sample of each loop. For example, for acidic FGF (B) the maximal values (in Å) are 5.1 (FF), 5.6 (ASPs), and 5.0 (bf) compared to the optimized values 2.4, 1.2, and 1.3, respectively. For RNase A, the corresponding maximal values are 4.1, 4.4, and 4.3 compared to the optimized values 1.1, 0.9, and 1; similar results have been obtained for the other loops.

It would be of interest to compare the best-fit RMSD values of Table II to those obtained in previous work. For loop 1 of the antibody, Bruccoleri et al.[45] obtained 2.6 (BB) and 3 Å (TOT) compared to 0.9 (BB) and 1.7 (TOT) Å obtained in Table III with ASP (bf). For loop 2 of the antibody they obtained 1.1 (BB) and 2.9 Å (TOT), Sudarsanam et al.[23] obtained 1.88 Å (BB), Collura et al.,[82] 0.93 Å (BB), compared to 0.6 (BB) and 0.8 Å (TOT) in Table III. For loop 3 of RNase we obtained in article I 0.20 (BB) and 1.86 (SC) compared to 0.2 (BB) and 1.4 Å (SC) in Table III. For myoglobin (119–125) Zheng and Kyle[83] obtained 0.16 (BB) and 0.35 (SC) compared to 0.2 (BB) and 0.5 Å (SC) in Table III. For loop 1 of BPTI, Sudarsanam et al. obtained

0.53 (BB), Collura et al. 0.92 (BB), Dudek and Scheraga 0.23 (BB), compared to 0.1 (BB) and 0.8 Å (SC) in Table III. For loop 2 of BPTI, Zheng and Kyle obtained 0.09 (BB) and 0.26 (SC), Zheng et al.[39] 0.84 (BB) and 2.33 Å (TOT), compared to 0.0 (BB), 0.9 (SC) and 0.7 Å (TOT) in Table III. Thus, the best-fit results in Table III are always comparable to or better than those obtained by the other methods besides for myoglobin where Zheng and Kyle studied a slightly different loop and for their RMSD (SC) for loop 2 of BPTI. As one would expect, for the highly stretched loops of myoglobin and BPTI all the methods have led to very small RMSD values.

TABLE IV. Differences in the Energy and the Free Energy Obtained by the LS Method†

Loop	$\Delta E^m_{\text{tot}}(n,\sigma_i^*)$	ΔE	ΔF
Acidic FGF	4.6	4.6	4.0
RNase A	1.8	0.9	1.1
Antibody, loop 1	4.9	4.6	4.3

†ΔE^m_{tot} (n,σ^*) (Eq. 2) is the difference between the minimized energy of NOS and the GEM. ΔE is the difference between the average energy of the wide microstates of NOS and GEM obtained from the MC simulations; ΔF is the corresponding difference in the free energy obtained by the convergence of approximate differences in the free energy (Eq. 6).

Free Energy Calculations

As discussed earlier, the minimized energy criterion for the loop stability is approximate and should be replaced by the free energy of wide microstates; the importance of including entropic effects in loop prediction algorithms has been emphasized recently by Honig's group.[84] Because lower energy microstates are typically characterized by smaller entropy, differences in free energy are expected to be smaller than the corresponding energy differences. Because of limitations in computer power (see below), the free energy has been studied only for three representative loops, loop 1 of the antibody (the largest loop), the loop of acidic FGF (a charged loop), and the loop of RNase A. Using an MC procedure in Cartesian coordinates, we carried out six runs starting from the minimized NOS and GEM structures of these loops based on the optimal ASPs; the free energy was calculated from these samples using the LS method. The MC runs were performed in the same way as described for the cyclic peptide in Ref. 47. Thus, at every MC step of each of the 3*N* Cartesian coordinates, the loop was randomly changed within ±0.008 Å (acidic FGF) and ±0.005 Å (the other two loops) around its current positions (i.e., within spheres of radii 0.014 and 0.009 Å, respectively). A trial conformation was accepted or rejected according to the usual Metropolis criterion[85] at temperature $T = 300$ K.

To avoid an "escape" of the structure from the original wide microstate, which is expected to occur during a long MC run, as in previous work,[47,49,86] the sample was created from shorter sets of runs each starting from the NOS (GEM) structure based on a different sequence of random numbers. The first 5000 MC steps of each set were ignored for relaxation and then every 100 MC steps the current conformation (dihedral and bond angles) was recorded for a later analysis, where 700 such structures define a set. The free energy was calculated from the entire sample of 28,000 loop structures obtained from 40 sets. It should be pointed out that the MC simulation of a loop is much more time consuming than of a cyclic peptide of a comparable size due to the need to calculate the loop-template interactions. Thus, generating an entire single sample for acidic FGF and RNase A required ~110 h CPU for each on a PC equipped with a 1.33 GH Athlon processor, whereas the entire MC sample of the larger loop 1 of the antibody required close to 400 h CPU. A well-known fact is that for proteins this MC procedure performs significantly worse than molecular dynamics (MD); however, MC leads to stable samples, while from our experience with cyclic peptides, such stability could not be obtained with MD. Therefore, we used here the MC method again but applied it only to three representative loops of 5, 9, and 12 residues. The stability of the samples was verified by methods described in detail in previous publications.[47,49]

Table IV presents results for the differences between the NOS and GEM of the minimized energies (gap energies), ΔE^m_{tot} (n,σ_i^*) (Eq. 2), the average MC energies, $\Delta E(n,\sigma_i^*)$, and the converged free energies, $\Delta F(n,\sigma_i^*)$ (Eq. 6) obtained with the LS method for the corresponding wide microstates. As expected, $\Delta F(n,\sigma_i^*)$ is smaller than ΔE^m_{tot} (n,σ_i^*), but only by 0.6–0.7 kcal/mol, which for acidic FGF and the antibody is still not smaller than the required 2 kcal/mol range. This provides a strong indication that the energy and free energy gaps should further be decreased by enhancing the screening of the electrostatic interactions [e.g., by increasing the parameter n (Eq. 1)], as discussed in the Summary.

Testing the Best-Fit Set of ASPs As Applied to an Independent Group of Loops

The best-fit set of ASPs has been determined as the set that does not differ significantly from the individual sets of ASPs optimized for the 12 "training" group of loops. Although the quality of the bf results for these loops has already been discussed, the performance of ASPs (bf) is checked further by calculating energy gaps and RMSD values based on this set and the force field alone for an independent "test" group of eight loops. For these loops, the templates have been defined with a radius of 9 Å, where only for the loop of RNase A a radius of 10 Å has been applied. Each set of results is based on an LTD search run of ~3000 structures starting from the NOS (as discussed earlier, this is not a limitation because a relatively large part of conformational space is searched). All these loops have B-factors < 40, most of them are not stretched, and they have been verified by computer graphics to be surface loops and not just protein segments. The proteins, the loops, and their sequences, together with the energy gaps, and the results for RMSD (BB) and RMSD (TOT) appear in Table V. The table reveals that the best-fit energy gaps are smaller than their FF counterparts besides for the last loop in the table (of the antibacterial

TABLE V. Energy Gaps and Heavy Atom RMSD From the NOS Structures Based on the Best-Fit Set of ASPs and $\varepsilon = 2r$, for an Independent Group of Eight Loops[†]

			FF			Best-fit ASPs		
			GAP	RMSD (Å)		GAP	RMSD (Å)	
Protein/loop	Sequence	R	(kcal/mol)	BB	TOT	(kcal/mol)	BB	TOT
RNase A (1rat) Loop1 (64–71) (8)	AC**K**NGQTN	3.2	0.6	0.4	0.9	0.4	0.4	0.9
Ser-Proteinase (2ptn) 143–151 (9)	NT**K**SSGTSY	4.9	6.9	2.1	2.4	3.9	0.6	0.6
Proteinase (2apr) 188–196 (9)	I**D**NS**R**GWWG	4.5	10.0	0.3	1.5	4.7	0.2	0.9
Proteinase (2apr) 128–137 (10)	**D**TITTV**R**GV**K**	4.3	14.8	1.3	2.3	3.3	0.8	1.0
Peptidase (5cpa) 244–250 (7)	ITTIYQA	2.7	9.0	0.7	1.3	3.4	0.6	2.2
RNase H (2rn2) 57–63 (7)	**E**AL**KEH**C	1.6	14.0	0.2	1.5	9.4	0.2	1.9
Antibody (1mcp) 56L–62L (7)	GAST**RE**S	1.3	9.8	0.1	0.7	8.8	0.7	1.1
Antibacterial pro. (1noa) 25–30 (6)	GLQAGT	1.3	0.5	0.1	0.6	1.7	0.1	0.7

[†]R is defined in the caption of Table I and the best-fit set of ASPs in Table II. The FF results are based on the force field with a distance-dependent dielectric constant $\varepsilon = 2r$. Charged residues in the sequences are set bold.

protein), where the energy gap increases from 0.5 (FF) to 1.7 kcal/mol (bf); however, this loop is relatively non-polar and stretched like the loop of myoglobin (Table II) for which a similar increase in the energy gap has been observed as well. The energy gaps of Table V are comparable to those obtained for the first group of loops in Table II. In addition; the relatively large gaps of 9.4 and 8.8 kcal/mol obtained for the seven residue loops of RNase H and the antibody (which consist of a relatively large number of charged residues), are in accord with the large gaps obtained for the charged loops of acidic FGF and AK in Table III (notice that although the charge of these loops and their templates is neutralized, the corresponding dipole moments are relatively strong).

The RMSD (bf) values decrease or actually remain unchanged compared to those of RMSD (FF) for the first four loops and the last one of Table V. For serine proteinase, a strong decrease from 2.1 Å (FF) to 0.6 Å (bf) is observed for RMSD (BB), which is similar to the corresponding decrease from 1.9 to 0.3 Å, obtained for proteinase in Table II; this finding shows that the effect of the ASPs on RMSD (BB) can be significant. For the loops of RNase H and the antibody, a moderate increase of 0.4 Å in RMSD (TOT) is observed, becoming 0.9 Å for peptidase in going from FF to bf. Notice that for RNase H and peptidase, RMSD (BB) actually remains the same where the increase in RMSD (bf,TOT) is due to the increase in RMSD (bf) of the side-chains. However, altogether for six or eight-loops the RMSD (bf,TOT) values in the table are close to or smaller than 1 Å.

SUMMARY

In this article, our general optimization procedure based on the solvation model of Eq. 1 has been tested as applied to 20 surface loops of different size and amino acid sequences from different proteins, which to the best of our knowledge, is the most extensive study carried out by the energetic approach to date. Because of its relative simplicity, the performance of this model could be studied, various computational issues have been investigated, and benchmark results have been provided for comparison with future calculations based on more sophisticated models.

We have studied for the first time in a systematic way the effect of loop flexibility (i.e., the ratio of the distance between the loop's ends and the loop length) and geometrical constraints imposed by the template on the optimization of the ASPs and have found, as expected, that the ASPs do not affect the highly stretched loops. We have also discussed the problems stemming from the uncertainty in the X-ray structures of surface loops. It has become evident that to be able to treat the highly charged loops, the electrostatic interactions should be weakened and thus as a first step we neutralized the charged amino acid residues. As in article I, we have found that for a typical hydrophilic loop that is attracted by the solvent and is free to protrude outward, all the ASPs, including that of carbon are negative, whereas for hydrophobic loops or loops with geometrical constraints imposed by the template, the ASP of carbon and other atoms can become positive.

Individual sets of ASPs were optimized for the 12 loops to examine whether a reasonable best-fit set can be defined from the individual sets (i.e., a set that does not differ significantly from each one of them). The best-fit set so defined was applied to the 12 loops as well as to a test group of 8 surface loops, where in most cases it has led to significantly smaller or comparable energy gaps and RMSD values than those obtained with the force field alone. This suggests that the best-fit set of ASPs can already be used to determine the structure of missing loops in homology modeling. Because in this case the correct loop structure is

unknown, the template will be defined with respect to an arbitrarily generated loop structure where the template's radius is increased, say from 10 to 12 Å.

However, the present results are not perfect, where the energy and free energy gaps, especially for the charged loops, are >3 kcal/mol. Decreasing these gaps would require the screening of the electrostatic interactions to be enhanced further by using larger values of the parameter *n,* for example; indeed, our calculations based on the force-field energy alone (E_{FF}) have shown that energy gaps obtained with $n = 3$ are smaller by up to 3 kcal/mol than those obtained with $n = 2$. However, the electrostatic energy prefers the collapsed loop structures, whereas the ASPs act in the opposite direction, meaning that an increased n would lead to less effective ASPs (i.e., with smaller absolute values). Because the electrostatic interaction is the most sensitive interaction to conformational changes, for larger n E_{tot} becomes less sensitive to such changes and the optimization procedure based typically on ~600 significantly different structures (according to the 60° criterion) cannot distinguish between the different ASP values; in other words, an optimal set cannot be determined because a wide range of ASPs lead to the same energy gap. Therefore, an optimization with $n > 2$ requires adopting a stricter criterion of variance (e.g., 40°), which, however, would lead to a significant increase in the number of conformations used in the optimization process, hence to an increase in the computation time, which is already large for $n = 2$ (see article I); this project will be carried out in the future, when enhanced computer capabilities become available for us.

Finally, it should be stressed again that our optimization procedure is general and can be used to optimize parameters of other implicit solvation models, such as the GB/SA model,[43,87–92] the screened Coulomb potential models of Mehler et al.,[93,94] the model of Lazaridis and Karplus,[95] the Poisson Boltzmann approach,[41] and the Protein Dipoles Langevin Dipoles (PDLD) approach of Warshel and collaborators.[96] Thus, the investigation of the present simplified model should be considered only as a first step, which has enabled testing various aspects of the methodology for surface loops, examining the effectiveness of the model as a predictor of loop structures, and generating a large number of benchmark results for comparison with future studies based on $n > 2$ and the above more sophisticated solvation models, where their performance as predictors of loop structures has not been studied in a systematic way thus far.

REFERENCES

1. Karplus PA, Schulz GE. Prediction of chain flexibility in proteins. Naturwissenschaften 1985;72:212–213.
2. Getzoff ED, Geysen HM, Rodda SJ, Alexander H, Tainer JA, Lerner RA. Mechanisms of antibody binding to a protein. Science 1987;235:1191–1196.
3. Rini JM, Schulze-Gahmen U, Wilson IA. Structural evidence for induced fit as a mechanism for antibody-antigen recognition. Science 1992;255:959–965.
4. Constantine KL, Friedrichs MS, Wittekind M, Jamil H, Chu CH, Parker RA, Goldfarb V, Mueller L, Farmer BT. Backbone and side chain dynamics of uncomplexed human adipocyte and muscle fatty acid-binding proteins. Biochemistry 1998;37:7965–7980.
5. Nicholson LK, Yamazaki T, Torchia DA, Grzesiek S, Bax A, Stahl SJ, Kaufman JD, Wingfield PT, Lam PYS, Jadhav PK, Hodge CN, Domaille PJ, Chang C-H. Flexibility and function in HIV-1 protease. Struct Biol 1995;2:274–280.
6. Collins JR, Burt SK, Erickson JW. Flap opening in HIV-1 protease simulated by "activated" molecular dynamics. Struct Biol 1995;2: 334–338.
7. Wagner G. The importance of being floppy. Struct Biol 1995;2:255–257.
8. Fetrow JS. Omega loops: nonregular secondary structures significant in protein function and stability. FASEB J 1995;9:708–717.
9. Crasto CJ, Feng J. Sequence codes for extended conformation: a neighbor-dependent sequence analysis of loops in proteins. Proteins 2001;42:399–413.
10. Leszcynski JF, Rose GD. Loops in globular proteins: a novel category of secondary structure. Science 1986;234:849–855.
11. Donate LE, Rufino SD, Canard LH, Blundell TL. Conformational analysis and clustering of short and medium size loops connecting regular secondary structures: a database for modeling and prediction. Protein Sci 1996;5:2600–2616.
12. Fechteler T, Dengler U, Schomburg D. Prediction of protein three-dimensional structures in insertion and deletion regions: a procedure for searching data bases of representative protein fragments using geometric scoring criteria. J Mol Biol 1995;253: 114–131.
13. Kwasigroch JM, Chomilier J, Mornon JP. A global taxonomy of loops in globular proteins. J Mol Biol 1996;259:855–872.
14. Martin AC, Toda K, Stirk HJ, Thornton JM. Long loops in proteins. Protein Eng 1995;8:1093–1101.
15. Oliva B, Bates PA, Querol E, Aviles FX, Sternberg MJ. An automated classification of the structure of protein loops. J Mol Biol 1997;266:814–830.
16. Ring CS, Kneller DG, Langridge R, Cohen FE. Taxonomy and conformational analysis of loops in proteins. J Mol Biol 1992;224: 685–699.
17. Bates PA, Sternberg MJ. Model building by comparison at CASP3: using expert knowledge and computer automation. Proteins 1999; Suppl 3:47–54.
18. Mosimann S, Meleshko R, James MN. A critical assessment of comparative molecular modeling of tertiary structures of proteins. Proteins 1995;23:301–317.
19. Sali A. Modeling mutations and homologous proteins. Curr Opin Biotechnol 1995;6:437–451.
20. Chothia C, Lesk AM. Canonical structures for the hypervariable regions of immunoglobulins. J Mol Biol 1987;196:901–917.
21. Chothia C, Lesk AM, Tramontano A, Levitt M, Smith-Gill SJ, Air G, Sheriff S, Padlan EA, Davies D, Tulip WR. Conformations of immunoglobulin hypervariable regions. Nature 1989;342:877–883.
22. Fidelis K, Stern PS, Bacon D, Moult J. Comparison of systematic search and database methods for constructing segments of protein structure. Protein Eng 1994;7:953–960.
23. Sudarsanam S, DuBose RF, March CJ, Srinivasan S. Modeling protein loops using a phi i + 1, psi i dimer database. Protein Sci 1995;4:1412–1420.
24. Summers NL, Karplus M. Modeling of globular proteins: a distance-based data search procedure for the construction of insertion/deletion regions and Pro–non-Pro mutations. J Mol Biol 1990;216: 991–1016.
25. Tappura K. Influence of rotational energy barriers to the conformational search of protein loops in molecular dynamics and ranking the conformations. Proteins 2001;44:167–179.
26. Deane CM, Blundell TL. A novel exhaustive search algorithm for predicting the conformation of polypeptide segments in proteins. Proteins 2000;40:135–144.
27. Bruccoleri RE, Karplus M. Prediction of the folding of short polypeptide segments by uniform conformational sampling. Biopolymers 1987;26:137–168.
28. Moult J, James MN. An algorithm for determining the conformation of polypeptide segments in proteins by systematic search. Proteins 1986;1:146–163.
29. van Vlijmèn HW, Karplus M. PDB-based protein loop prediction: parameters for selection and methods for optimization. J Mol Biol 1997;267:975–1001.
30. Wojcik J, Mornon JP, Chomilier J. New efficient statistical sequence-dependent structure prediction of short to medium-sized

protein loops based on an exhaustive loop classification. J Mol Biol 1999;289:1469–1490.
31. Das B, Meirovitch H. Optimization of solvation models for predicting the structure of surface loops in proteins. Proteins 2001;43:303–314.
32. Fine RM, Wang H, Shenkin PS, Yarmush DL, Levinthal C. Predicting antibody hypervariable loop conformations. II. Minimization and molecular dynamics studies of MCPC603 from many randomly generated loop conformations. Proteins 1986;1:342–362.
33. Higo J, Collura V, Garnier J. Development of an extended simulated annealing method: application to the modeling of complementary determining regions of immunoglobulins. Biopolymers 1992;32:33–43.
34. Rosenfeld R, Zheng Q, Vajda S, DeLisi C. Computing the structure of bound peptides. Application to antigen recognition by class I major histocompatibility complex receptors. J Mol Biol 1993;234:515–521.
35. Shenkin PS, Yarmush DL, Fine RM, Wang HJ, Levinthal C. Predicting antibody hypervariable loop conformation. I. Ensembles of random conformations for ringlike structures. Biopolymers 1987;26:2053–2085.
36. Caralacci L, Englander SW. Loop problem in proteins: developments on the Monte Carlo simulated annealing approach. J Comput Chem 1996;17:1002–1012.
37. Dudek MJ, Scheraga HA. Protein structure prediction using a combination of sequence homology and global energy minimization. I. Global energy minimization of surface loops. J Comput Chem 1990;11:121–151.
38. Gō N, Scheraga HA. Ring closure and local conformational deformations of chain molecules. Macromolecules 1970;3:178–187.
39. Zheng Q, Rosenfeld R, Vajda S, DeLisi C. Loop closure via bond scaling and relaxation. J Comput Chem 1993;14:556–565.
40. Mas MT, Smith KC, Yarmush DL, Aisaka K, Fine RM. Modeling the anti-CEA antibody combining site by homology and conformational search. Proteins 1992;14:483–498.
41. Smith KC, Honig B. Evaluation of the conformational free energies of loops in proteins. Proteins 1994;18:119–132.
42. Wesson L, Eisenberg D. Atomic solvation parameters applied to molecular dynamics of proteins in solution. Protein Sci 1992;1:227–235.
43. Qiu D, Shenkin PS, Hollinger FP, Still WC. The GB/SA continuum model for solvation. A fast analytical method for the calculation of approximate Born radii. J Phys Chem 1997;101:3005–3014.
44. Rapp CS, Friesner RA. Prediction of loop geometries using generalized Born model of solvation effects. Proteins 1999;35:173–183.
45. Bruccoleri RE, Haber E, Novotny J. Structure of antibody hypervariable loops reproduced by a conformational search algorithm. Nature 1988;335:564–568.
46. Baysal C, Meirovitch H. Determination of the stable microstates of a peptide from NOE distance constraints and optimization of atomic solvation parameter. J Am Chem Soc 1998;120:800–812.
47. Baysal C, Meirovitch H. Free energy based populations of interconverting microstates of a cyclic peptide lead to the experimental NMR data. Biopolymers 1999;50:329–344.
48. Meirovitch H, Meirovitch E, Lee J. New theoretical methodology for elucidating the solution structure of peptides from NMR data. I. The relative contribution of low energy microstates to the partition function. J Phys Chem 1995;99:4847–4854.
49. Meirovitch H, Meirovitch E. New theoretical methodology for elucidating the solution structure of peptides from NMR data. 3. Solvation effects. J Phys Chem 1996;100:5123–5133.
50. Baysal C, Meirovitch H. On the transferability of atomic solvation parameters: Ab initio structural prediction of cyclic heptapeptides in DMSO. Biopolymers 2000;54:416–428.
51. Baysal C, Meirovitch H. Ab initio prediction of the solution structures and populations of a cyclic pentapeptide in DMSO based on an implicit solvation model. Biopolymers 2000;53:423–433.
52. Baysal C, Meirovitch H. Efficiency of the local torsional deformations method for identifying the stable structures of cyclic molecules. J Phys Chem 1997;101:2185–2191.
53. Baysal C, Meirovitch H. Efficiency of simulated annealing for peptides with increasing geometrical constraints. J Comput Chem 1999;20:1659–1670.
54. Meirovitch H. Calculation of entropy with computer simulation methods. Chem Phys Lett 1977;45:389–392.
55. Meirovitch H, Koerber SC, Rivier JE, Hagler AT. Computer simulation of the free energy of peptides with the local states method: analogues of gonadotropin releasing hormone in the random coil and stable states. Biopolymers 1994;34:815–839.
56. Ponder JW. TINKER-software tools for molecular design. St. Louis:Washington University 1999; Version 3.7.
57. Jorgensen WL, Maxwell DS, Tirado-Rives J. Development and testing the OPLS all-atom force field on conformational energetics and properties of organic liquids. J Am Chem Soc 1996;118:11225–11236.
58. MacKerell AD Jr, Bashford D, Bellott M, Dunbrack RL Jr, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph McCarthy D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnik S, Ngo T, Nguyen DT, Prodhom B, Reiher WE III, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiórkiewicz-Kuczera J, Yin D, Karplus M. All-atom empirical potential for molecular modeling and dynamic studies of proteins. J Phys Chem 1998;102:3586–3616.
59. Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM Jr, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. J Am Chem Soc 1995;117:5179–5197.
60. Santoro J, Gonzalez C, Bruix M, Neira JL, Nieto JL, Herranz J, Rico M. High-resolution three-dimensional structure of ribonuclease A in solution by nuclear magnetic resonance spectroscopy. J Mol Biol 1993;229:722–734.
61. Kuriyan J, Osapay K, Burley SK, Brunger AT, Hendrickson WA, Karplus M. Exploration of disorder in protein structures by X-ray restrained molecular dynamics. Proteins 1991;10:340–358.
62. Svenson LA, Sjolin L, Dill J, Gilliand GL. The conformation flexibility of surface residues of bovine pancreatic ribonuclease A at 1.1Å. In: Cuchillo CM, de Llorens R, Nogues MV, Pares X, editors. Proceedings of the 2nd International Meeting, Universitat Autonoma de Barcelona, Barcelona 1991;31–38.
63. Tilton RF Jr., Dewan JC, Petsko GA. Effects of temperature on protein structure and dynamics: X-ray crystallographic studies of the protein ribonuclease-A at nine different temperatures from 98 to 320 K. Biochemistry 1992;31:2469–2481.
64. Elber R, Karplus M. Multiple conformational states of proteins-a molecular dynamics analysis of myoglobin. Science 1987;235:318–321.
65. Karplus M, Kushik JN. Method for estimating configurational entropy of macromolecules. Macromolecules 1981;14:325–332.
66. Stillinger FH, Weber TA. Packing structures and transitions in liquids and solids. Science 1984;225:983–989.
67. Meirovitch H. Calculation of the free energy and entropy of macromolecular systems by computer simulation. In: Lipkowitz KB, Boyd DB, editors. Rev Comp Chem 1998;12:1–74.
68. Beveridge DL, DiCapua FM. Free energy via molecular simulation: applications to chemical and biomolecular systems. Annu Rev Biophys Biophys Chem 1989;18:431–492.
69. Tanner JJ, Nell LJ, McCammon JA. Anti-insulin antibody structure and conformation. II. Molecular dynamics with explicit solvent. Biopolymers 1992;32:23–32.
70. Ooi T, Oobatake M, Némethy G, Scheraga HA. Accessible surface areas as a measure of the thermodynamic parameters of hydration of peptides. Proc Natl Acad Sci USA 1987;84:3086–3090.
71. Schiffer CA, Caldwell JW, Kollman PA, Stroud RM. Protein structure prediction with a combined solvation free energy-molecular mechanics force field. Mol Simul 1993;10:121–149.
72. Fraternali F, Van Gunsteren WF. An efficient mean solvation force model for use in molecular dynamics simulations of proteins in aqueous solution. J Mol Biol 1996;256:939–948.
73. Gilson MK, Honig B. The inclusion of electrostatic hydration energies in molecular mechanics calculations. J Comput Aided Mol Des 1991;5:5–20.
74. Najmanovich R, Kuttner J, Sobolev V, Edelman M. Side-chain flexibility in proteins upon ligand binding. Proteins 2000;39:261–268.
75. Zhao S, Goodsell DS, Olson AJ. Analysis of a data set of paired uncomplexed protein structures: new metrics for side-chain flexibility and model evaluation. Proteins 2001;43:271–279.
76. Wilson MA, Brunger AT. The 1.0 A crystal structure of Ca(2+)-bound calmodulin: an analysis of disorder and implications for functionally relevant plasticity. J Mol Biol 2000;301:1237–1256.
77. Esposito L, Vitagliano L, Sica F, Sorrentino G, Zagari A, Mazza-

rella L. The ultrahigh resolution crystal structure of ribonuclease A containing an isoaspartyl residue: hydration and sterochemical analysis. J Mol Biol 2000;297:713–732.
78. Muller CW, Schulz GE. Structure of the complex between adenylate kinase from Escherichia coli and the inhibitor Ap5A refined at 1.9 A resolution. A model for a catalytic transition state. J Mol Biol 1992;224:159–177.
79. Flory PJ. Statistical mechanics of chain molecules. New York: Hasner; 1988.
80. Liu DC, Nocedal J. On the limited memory BFGS method for large scale optimization. Technical Report NAM03, Evanston, IL, Department of Electrical Engineering and Computer Science, North Western University; 1988.
81. Perrot G, Cheng B, Gibson KD, Palmer KA, Nayeem A, Maigret B, Scheraga HA. MSEED—a program for rapid analytical determination of of accessible surface area s and their derivatives. J Comput Chem 1992;13:1–11.
82. Collura V, Higo J, Garnier J. Modeling of protein loops by simulated annealing. Protein Sci 1993;2:1502–1510.
83. Zheng Q, Kyle DJ. Accuracy and reliability of the scaling-relaxation method for loop closure: an evaluation based on extensive and multiple copy conformational samplings. Proteins 1996; 24:209–217.
84. Xiang X, Soto CS, Honig B. Evaluating conformational free energies: the colony energy and its application to the problem of loop prediction. Proc Natl Acad Sci USA 2002;99:7432–7437.
85. Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E. Equation of state calculations by fast computing machines. J Chem Phys 1953;21:1087–1092.
86. Brady J, Karplus M. Conformational entropy of the alanine dipeptide in vacuum and in solution: a molecular dynamics study. J Am Chem Soc 1985;107:6103–6105.
87. Haukins GD, Liotard DA, Cramer CJ, Truhlar DG. OMNISOL: fast prediction of free energy of solvation and partition coeeficients. J Org Chem 1998;63:4305–4313.
88. Calimet N, Schaefer M, Simonson T. Protein molecular dynamics with the generalized Born/ACE solvent model. Proteins 2001;45: 144–158.
89. Gallicchio E, Zhang LY, Levy RM. The SGB/NP hydration free energy model based on the surface generalized born solvent reaction field and novel nonpolar hydration free energy estimators. J Comput Chem 2002;23:517–529.
90. Schaefer M, Karplus M. A comprehensive analytical treatment of continuum electrostatics. J Phys Chem 1996;100:1578–1599.
91. Zhang LY, Gallicchio E, Friesner RA, Levy RM. Solvent models for protein-ligand binding: comparison of implicit solvent Poisson and surface generalized born models with explicit solvent simulations. J Comput Chem 2001;22:591–607.
92. Dominy BN, Brooks CI III. Development of a generalized Born model parametrization for proteins and nucleic acids. J Phys Chem 1999;103:3765–3773.
93. Hassan SA, Guarnieri F, Mehler EL. A general treatment of solvent effects based on screened Coulomb potentials. J Phys Chem B 2000;104:6478–6489.
94. Hassan SA, Mehler EL. A critical analysis of continuum electrostatics: the screened Coulomb potential-implicit solvent model and the study of the alanine dipeptide and discrimination of misfolded structures of proteins. Proteins 2002;47:45–61.
95. Lazaridis T, Karplus M. Effective energy function for proteins in solution. Proteins 1999;35:133–152.
96. Warshel A, Russell ST. Calculation of electrostatic interactions in biological systems and in solutions. Q Rev Biophys 1984;17:283–422.