JCTC Journal of Chemical Theory and Computation

Optimized Radii for Poisson−Boltzmann Calculations with the AMBER Force Field

Jessica M. J. Swanson,*[,†,‡,§] Stewart A. Adcock,[†,‡,§] and J. Andrew McCammon[†,‡,§,||]

Howard Hughes Medical Institute, Center for Theoretical Biological Physics, Department of Chemistry and Biochemistry, and Department of Pharmacology, University of California at San Diego, La Jolla, California 92093-0365

Received December 29, 2004

Abstract: Implicit solvent models are a standard tool for assessing the electrostatics of biomolecular systems. The accuracy of quantitative predictions, such as pK_a values, transfer free energies, binding energies, and solvation forces, is strongly dependent on one's choice of continuum parameters: the solute charges, dielectric coefficient, and radii, which define the dielectric boundary. To ensure quantitative accuracy, these parameters can be benchmarked against explicit solvent simulations. Here we present two sets of optimized radii to define either abrupt or cubic-spline smoothed dielectric boundaries in Poisson−Boltzmann calculations of protein systems with AMBER (parm99) charges. Spline smoothing stabilizes the electrostatic potential at the molecular surface, allowing for continuum force calculations. Most implementations, however, require significantly different radii than the abrupt boundary surfaces. The optimal continuum radii are initially approximated from the solvent radial charge distribution surrounding each atom type. A genetic algorithm is then used to fine-tune the starting values to reproduce charging free energies measured from explicit solvent simulations. The optimized radii are tested on four protein-like polypeptides. The results show increased accuracy of molecular solvation energies and atomic forces relative to commonly used continuum parameter sets. These radii are suitable for Poisson−Boltzmann calculations with the AMBER force field and offer energetic congruence to any model that combines molecular mechanics and Poisson−Boltzmann solvation energies.

I. Introduction

Continuum solvation models are a useful link between the microscopic and macroscopic realms of theoretical biochemistry and are widely used to evaluate the electrostatics of biomolecular systems. One of the most common models solves the Poisson Equation (PE)[1] or the Poisson−Boltzmann Equation (PBE)[2,3] as a function of the solute's charge density, a spatially dependent dielectric coefficient, and, in the case of the PBE, ion concentration. The resulting electrostatic potential can be used, for example, to predict electrostatic complementarity of molecular surfaces, to estimate electrostatic binding and transfer free energies, to approximate pK_a values, and to supply electrostatic forces for use in molecular dynamics.

The accuracy of continuum models is dependent on the parameters that are used to define the solute charges, the solvent and solute dielectric coefficients, and the atomic radii which define the dielectric boundary. It has been shown that the average solvent charge density in the continuum model is a sharply peaked function in the region of dielectric discontinuity.[4] The solvent charge density gives rise to the solvent reaction field. Thus, quantitative results are especially sensitive to the location of the solvent−solute dielectric

* Corresponding author phone: (858)822-2771; fax: (858)534-4974; e-mail: jswanson@ucsd.edu.
† Howard Hughes Medical Institute.
‡ Center for Theoretical Biological Physics.
§ Department of Chemistry and Biochemistry.
|| Department of Pharmacology.

10.1021/ct049834o CCC: $30.25
Published on Web 03/31/2005

boundary, i.e., the molecular surface. To appreciate where the dielectric boundary should be, it is helpful to relate its macroscopic and microscopic descriptions. Early studies comparing atomistic simulations of aqueous solutions of monatomic ions to the Born model were helpful in elucidating this connection.[5−7] These studies revealed significant differences in the solvent structure surrounding anions and cations with the same ionic radius but different solvation energies. The Born model correctly predicts disparate anionic and cationic solvation energies if, and only if, the radii used to define the continuum dielectric boundary correspond with the first peak in the microscopic solvent density. Thus, the optimal Born radii are not the ionic radii, a function of the ion alone, but those that trace out solvent excluded cavities, which are a function of both the ion and the surrounding solvent structure.

Optimal biomolecular radii could, presumably, be measured in a similar manner. Biomolecules are, however, more complicated because the solvent structure surrounding each atom is influenced by neighboring atoms and the solute's conformation. Continuum radii for biomolecules must, therefore, balance accuracy for a given conformation with robustness across multiple conformations. Though the solvent structure can still be used to approximate the location of the dielectric boundary, the radii that define that boundary must ultimately be benchmarked against quantitative observables such as solvation energies or forces. Several continuum parameter sets, consisting of solute charges and radii that complement a specified protein dielectric constant and molecular surface definition, have been optimized to reproduce either experimental solvation energies[8,9] or explicit solvent simulations.[10,11] These two methods will be considered in turn.

Referencing continuum models against experimentally determined solvation free energies is an appealing approach because there are many small molecule data to draw upon and because computational models should ultimately be benchmarked by experiment. There are, however, two disadvantages to this approach. First, it necessitates estimating the nonpolar contribution to the solvation free energy since electrostatic and nonelectrostatic contributions cannot be clearly distinguished in experiment. Solvation energies, rigorously defined as the reversible work involved in transferring a molecule from gas phase to bulk solvent, are typically separated into electrostatic and nonelectrostatic contributions via a thermodynamic cycle. This cycle involves cavity formation, the introduction of solute−solvent vdW interactions, and the introduction of solute−solvent electrostatic interactions; $\Delta G_{solv} = \Delta G_{cavity} + \Delta G_{vdW} + \Delta G_{elec}$. PB calculations only account for the electrostatic contribution, ΔG_{elec}. The remaining nonpolar contribution is commonly approximated by the Solvent Accessible Surface Area (SASA) model, $\Delta G_{np} = \Delta G_{cavity} + \Delta G_{vdw} \cong \gamma\ \Delta SASA$, despite its inadequacy as discussed by Gallicchio et al.[12] The second disadvantage is that one is limited to the small, neutral molecules for which experimental results are available. Accordingly, the transferability of these parameters to biomolecular systems such as proteins, where backbone hydrogen bonding and charged residues likely affect solvent−solute interactions, is unknown.

Benchmarking continuum parameters against explicit solvent simulations avoids the aforementioned challenges; simulations can be applied to charged molecules that mimic biomolecular systems, and thermodynamic cycles can be used to separate solvation energies into electrostatic and nonelectrostatic contributions. In the first leg of the thermodynamic cycle, the 'growth' phase, the nonpolar contributions are accounted for by growing a neutral solute cavity into bulk solvent and introducing solute−solvent vdW interactions. In the next leg, the 'charging' phase, the electrostatics contributions are measured by introducing solvent−solute charge−charge interactions. The main disadvantage to using explicit solvent simulations is the error inherent in modern force fields, limited sampling, and simulations techniques. It is encouraging that attempts to minimize these errors have become increasingly successful.[13,14] Simulations additionally offer energetic congruence to methods that combine continuum solvation and molecular mechanics energies such as end-point free energy calculations,[15,16] pK_a calculations[17,18] continuum dynamics,[19] and constant pH molecular dynamics.[20]

It is also possible to compare the forces acting on specific atoms from explicit solvent simulations and continuum models.[21] Continuum force evaluations require a smooth dielectric boundary because abrupt boundaries often result in numerical instability in the electrostatic potential. Techniques such as Gaussian[22] and cubic spline[23] based volume functions have been introduced to smooth the dielectric boundary, making continuum force calculations and continuum dynamics possible. Spline smoothed surfaces are now standard options in several Poisson−Boltzmann solvers.[24,25] To maintain quantitative accuracy, however, these implementations require significant rescaling of the radii used to define the dielectric boundary. If the radii are not rescaled, solvation energies and forces tend to be overestimated by 10 to 40% (results not shown). Nina et al. used explicit solvent simulations to optimize of a set of radii for the CHARMM22 force field for abrupt boundary definitions[10] and later rescaled these radii for spline smoothed boundary definitions.[10,11] Since the location of the dielectric boundary is highly dependent on the solute charge distribution, the Nina et al. radii are not transferable to other solute charge definitions. Thus, a similar effort is needed for other force fields as has been suggested by a number of authors.[19,21,26]

We present two sets of optimized radii for the AMBER (parm99) force field for either abrupt or cubic spline smoothed dielectric boundary definitions. The radii are initially approximated from the solvent charge distributions measured during explicit solvent simulations and then optimized with a genetic algorithm (GA) to reproduce explicit solvent charging energies. We present the relative performance of several commonly used continuum parameter sets on the model compounds and four protein-like polypeptide chains. The latter were included to demonstrate the transferability of these parameters to protein systems. Both sets of optimized radii improve the accuracy of continuum solvation energies and the smooth boundary radii improve

Table 1. 14 Polyalanine Peptides Used To Optimize the Backbone Radii

polypeptide[a]	residues	description[b]	H-bonding
l-beta1	17−20	type II	yes
l-beta2	36−39	type I′	yes
l-beta3	39−42	type I	yes
l-beta4	59−62	type IV	no
l-beta5	60−63	type I	no
l-beta6	85−88	type VIII	no
l-helix1	5−14	type H	yes
l-helix2	109−114	type H	yes
l-hairpin1	42−53	class 3:5 IG	yes
l-hairpin2	51−59	class 4:4	yes
c-beta1	17−20	type I	yes
c-beta2	42−45	type IV	no
c-helix1	7−20	type H	yes
c-helix2	23−30	type H	yes

[a] Peptides taken from lysozyme (1aki) and crambin (1ejg). [b] Secondary structure descriptions provided by PDBsum.[29]

the correlation between explicit and implicit forces. These radii are suitable for PB quantitative measurements with AMBER partial charges and are recommended for methods that combine AMBER molecular mechanics and PB solvation energies.

The next section describes the model systems and the methodology used in the explicit solvent simulations, the continuum calculations, and the genetic algorithm optimizations. Section 3 defines the radius groups, discusses the explicit solvent charging energies, compares continuum solvation energies from different parameter sets, verifies radii transferability to proteins with protein test cases, and presents the quantitative affects of these radii on atomic forces. Section 4 summarizes the work and possible future directions.

2. Methods

2.1. Model System. The explicit solvent simulations, continuum calculations, and genetic algorithm optimizations were divided into two stages. First, the protein backbone atoms were optimized using 14 polyalanine peptides of varying lengths in common secondary structure conformations. Each conformation (see Table 1) was modeled from fragments of either lysozyme (pdb code 1ati) or crambin (pdb code 1ejg). The fragments were mutated to polyalanine and terminated with neutral blocking groups with the MMTSB Tool Set.[27] Second, the side chain radii were optimized using 20 nonzwiterionic *N*-acetyl-X-*N*′-methylamide dipeptides where X represents one of the twenty standard amino acids. Two conformations of each side chain dipeptide were used. The first conformation, chosen for the sake of comparison with previous optimizations,[10] used extended backbone phi and psi angles (180°, 180°) and the most frequent side chain dihedral angles from a Dunbrack backbone-*independent* rotamer library. The second conformation used a much more common backbone conformation (−60°, −40°) and the most frequent rotamers from a Dunbrack backbone-*dependent* rotamer library.[28] All model conformations are provided in the Supporting Information.

2.2. Explicit Solvent Simulations. The AMBER parm99 force field converted to CHARMM format was used in all simulations. Hydrogen atoms were first energy minimized in vacuum with 50 steps of steepest descent followed by 1000 steps of the Adopted Basis Newton Raphson (ABNR) method. All solute atoms were then fixed for the duration of the simulation. Each model compound was solvated in a sphere of explicit TIP3P water molecules that extended 6.5 Å beyond the dipeptides and 10.0 Å beyond the polyalanine peptides. This resulted in 3−4 hydration shells around every solute atom. Running simulations with larger and smaller solvent shells verified that the chosen dimensions were sufficient for energetic convergence. The spherical solvent boundary potential (SSBP) model including Kirkwood's mulitpolar expansion reaction field was used to approximate the influence of bulk water beyond the explicit water sphere.[30] This model alleviates many of the difficulties that result from perturbing charged systems with periodic boundary conditions and has been shown to give reliable results for proteins, nucleic acids, and, most recently, small molecules.[4,14,31] The solvent was first energy minimized with 50 steps of steepest descent followed by 1000 steps of the ABNR method and then equilibrated for 100 ps.

All simulations employed Langevin dynamics at constant temperature (300 K) using SHAKE enabled 2 fs time steps, infinite cutoffs for nonbonded interactions, and a friction constant corresponding to a relaxation time of 5 ps applied to water oxygen atoms. The preequilibrated fully charged systems were simulated for 200 ps to obtain the solvent charge distribution surrounding each solute atom. The solvent charge distributions were used to verify that the continuum radii were properly grouped and to estimate their starting values.

The charging free energies were measured with free energy perturbation (FEP) simulations run in the PERT module of CHARMM.[25] Each simulation consisted of 10 windows in which the solute's charge was scaled by a thermodynamic coupling parameter λ varying by $\pm$ 0.1 from 0 to 1 and then from 1 to 0 according to, $q(\lambda) = \lambda q_{final}$. The weighted histogram analysis method (WHAM) was used to combine the results of the individual windows to calculate the total charging free energy. WHAM is a self-consistent iterative procedure that optimizes the distribution of data from separate simulations and thus decreases the amount of sampling required for convergence.[32] Two tests were used to ensure that the simulations were converged. First, the standard error, calculated as half the difference between the forward and reverse WHAM postprocessed results and shown in Table 3, was required to be less than 3% of the total energy. Second, simulations of twice the length were required to be within 2% of the original simulations for the largest and most charged model compounds (results not shown). Different window lengths were required for convergence by the backbone and side chain model compounds; the polyalanine peptides were equilibrated for 10 ps followed by 40 ps of collection, while the dipeptides showed convergence in 5 ps of equilibration followed by 20 ps of collection.

2.3. Continuum Calculations. All continuum calculations were performed with the Adaptive Poisson−Boltzmann Solver (APBS)[24] using zero bulk ionic strength, a temperature of 300 K, a solvent dielectric of 78.4, and a solute dielectric

Table 2. Radius Groups with GA Starting Values and the Final Optimized Values

atom name[a]	residues	start value	final value[b]	final value[c]
	Backbone			
C	all	2.30	1.903	2.170
O	all	1.58	1.454	1.742
N	all	2.50	1.835	2.262
CA	all except G	2.80	1.591	2.339
CA	G	2.60	1.654	2.133
CAY/CAT	ACE,NME	2.51	2.595	2.375
	Side Chains			
CB	D,E,C,H,M,F,S,T,W,Y	2.55	2.087	2.370
CB	A,R,N,Q,I,L,K,V	2.75	1.829	2.063
CG*	R,Q,I,L,K,M,T,V	2.49	2.039	2.329
CG	H,F,W,Y	2.10	1.651	2.290
CG/CD	N,Q,D	2.19	1.995	2.257
CG	E	2.45	1.942	2.432
CB/CG/CD	P	2.70	2.008	2.157
CD	R,K	2.81	2.034	2.303
CD*	I,L	2.45	1.897	2.103
CD*/CE*/CZ	H,F,W,Y	2.05	1.837	2.122
CE	M	2.40	1.902	2.157
CZ/CE	R,K	2.66	2.020	2.414
OD*/OE*	N,Q,D,E	1.55	1.516	1.727
OG*	S,T	1.65	1.562	1.832
OH	Y	1.72	1.738	2.022
NE,NH*,NZ	R,K	2.48	1.523	1.861
ND2/NE2	N,Q	2.12	2.222	2.453
ND1,NE2	H	1.90	1.436	1.782
NE1	W	2.11	1.898	2.147
SG/SD	C,M	2.00	1.978	2.169
	Hydrogens[d]			
type H	bb HN	1.20	1.600	1.967
type H	bound to N	1.20	1.119	1.379
type HO/HS	bound to O/S	1.00	1.201	1.406
type H1/HP	polar	1.31	1.914	2.033
type HC/HA	nonpolar	1.30	0.840	1.321

[a] Radius groups are distinguished by AMBER atom names for all heavy atoms and by atom type for hydrogen atoms. [b] Final radii for abrupt dielectric surfaces. [c] Final radii for spline smoothed surfaces. [d] Hydrogens specified by atom type with type 'H' divided into two groups; amide backbone 'HN' and all other N-bound hydrogens.

Table 3. FEP Charging and Continuum Solvation Energies for the 14 Polyalanine Peptides Using AMBER and Optimized Radii[c]

polypep	WHAM		AMBER	opt[a]	opt[b]
l-beta1	−34.17	(0.06)	−31.87	−35.44	−33.53
l-beta2	−32.69	(0.05)	−25.67	−30.68	−31.85
l-beta3	−30.89	(0.48)	−27.88	−31.40	−30.79
l-beta4	−31.42	(0.31)	−28.89	−32.26	−32.88
l-beta5	−33.69	(0.03)	−32.18	−34.83	−32.87
l-beta6	−31.09	(0.06)	−27.27	−31.53	−32.34
l-helix1	−65.29	(0.45)	−53.78	−64.15	−65.31
l-helix2	−49.18	(0.42)	−43.36	−48.18	−51.49
l-hairpin1	−66.50	(1.33)	−55.70	−66.14	−66.63
l-hairpin2	−81.45	(1.06)	−69.17	−81.60	−81.43
c-beta1	−34.72	(0.24)	−33.64	−34.54	−33.90
c-beta2	−40.90	(0.24)	−36.47	−40.88	−41.06
c-helix1	−68.91	(0.44)	−54.99	−66.39	−67.15
c-helix2	−51.91	(0.04)	−44.01	−51.05	−51.77

[a] Abrupt smoothed dielectric boundary definitions. [b] Spline smoothed dielectric boundary definitions. [c] All energies are in kcal/mol. Standard errors are reported as half the difference between the forward and reverse WHAM energies.

of 1. The PB grid was centered on each solute and extended at least 20 Å beyond its dimensions. A grid resolution of 0.25 Å was proven sufficient for energetic convergence by comparing calculations with 0.15 Å grid resolution, which resulted in a correlation coefficient of 0.9995 and an average absolute error (AAE) of 0.97 kcal/mol. Solute charges were distributed onto grid points using a cubic B-spline discretization. Electrostatic solvation free energies were calculated from the energetic difference in the solvated ($\epsilon_{bulk} = 78.4$) and gas-phase systems ($\epsilon_{bulk} = 1$). The molecular surface was defined by the interface of a 1.4 Å solvent probe and the solute radii.

Some of the choices made in the continuum calculations protocol warrant explanation. First, the molecular surface traced by rolling a solvent probe, $r = 1.4$ Å, around the solute atoms was used to define the location of the abrupt dielectric boundary. Previous PB radii optimizations[10] have used the van der Waals (vdW) surface made up of overlapping solute atoms which excludes the interstitial spaces that lie between solute crevices from the molecular volume. Although the vdW surface definition works well for small molecules, it often results in buried high dielectric pockets in larger molecules such as proteins.[33] These pockets can change quantitative results significantly, making the radii less robust across similar conformations. The vdW surface is particularly problematic in continuum dynamics as high dielectric pockets can appear and disappear as frequently as every time step, resulting in numerical instability.[19] Second, for the spline smoothed surfaces, cubic B-spline charge discretization was used instead of trilinear interpolation to avoid large orientational artifacts.[22] Finally, a solute dielectric of 1 was chosen for consistency with the nonpolarizable force field and fixed solute conformation.

2.4. GA Optimization. The GA was used to optimize the radii to their final values. The GA is an efficient stochastic optimization method that has been widely applied to minimization problems because it is ideally suited for multiple-dimensional global search problems where the search space contains multiple local minima and the search variables may or may not be correlated. The GA begins with the generation of an initial population of a given number of solutions. The fitness of each solution is evaluated, and a new population is generated via selection, crossover, and mutation. This process is repeated until a desired fitness is reached or the maximum number of generations exceeded.

In the initial evolutions, populations of 50 solutions were run for 100 generations. In subsequent evolutions, populations of 100 solutions run for 50 generations. Solutions in the initial population of a given evolution were generated from a uniform distribution ± 0.10 Å around the starting radii. In subsequent generations, solutions were generated via a process of selection followed by crossover or mutation. Selections were performed with the Stochastic Universal Sampling algorithm,[34] which is designed to give zero bias

in the selection and minimal spread. It selects solutions with a probability proportional to their fitness

$$p_i = \frac{F_i}{\sum_{j=1}^{N} F_j} \quad (1)$$

where p_i is the probability that solution i will be selected and F_i is its fitness. The fitness function was normalized to ensure that fitness scores remained between 0 (poor) and 1 (perfect)

$$F_i = \frac{1}{(1 - AAE)_i} \quad (2)$$

where AAE_i is the average absolute error of solution i. A uniform crossover process was applied to 90% of the population. Each crossover event consisted of randomly distributing the radii from two previous-generation solutions to two new-generation solutions. Mutations were applied to 20% of the population and consisted of perturbing a solution's radii using a Gaussian distribution with a standard deviation of ±0.05 Å. The evolution was terminated if the total fitness and the best solution converged to the same value for 6 consecutive generations. In the absence of convergence a new evolution was started with the best radii from the previous evolution.

3. Results and Discussion

3.1. Radii Grouping and Starting Values. In optimization procedures, the number of parameters that can be meaningfully optimized is generally limited by the number of reference values. Although there is likely an ideal radius for every atom in every conformation, it is desirable to find a set of radii that are robust across multiple conformations. It is generally accepted that atoms in similar chemical environments have comparable surrounding solvent structures and thus similar optimal continuum radii. The factors directly influencing solvent structure are the atom's charge, vdW parameters, and structural neighbors. It may seem appealing to use AMBER atom types to define a set of continuum radius groups. This does not work, however, since AMBER atom types are distinguished by vdW parameters and often have significantly different charges in different residues. Instead, atoms were initially grouped according to similar chemical environments. The groups were then tested for similar surrounding solvent structure.

The solvent charge distribution surrounding each atom type was measured from 100 ps of Langevin dynamics of solvent plus rigid, fully charged solute molecules. The first peak in the solvent charge distribution corresponds to the closest explicit water molecules and can be used to approximate the optimal continuum radius. Figure 1a shows almost identical solvent charge distributions for the carboxyl oxygen in both conformations of the aspartate and glutamate residues. As expected, the first peak is positive due to electropositive water hydrogens crowding around the negatively charged carboxyl oxygen. Conversely, the first peak in the solvent charge distribution surrounding positively charged atoms,

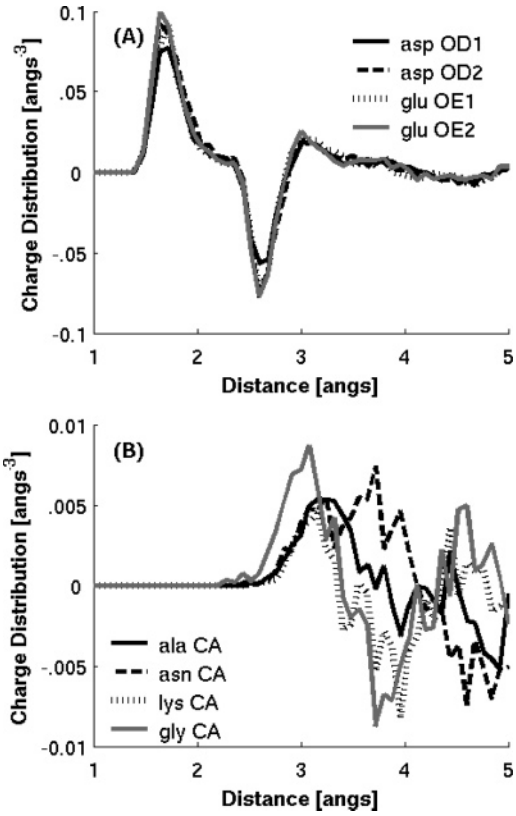

Figure 1. Radial solvent charge distributions show that similar chemical environments result in similar solvent structures around (A) the electronegative carboxyl oxygen in asp and glu as well as (B) the alpha carbons of ala, asn, and lys. The alpha carbon of gly reveals a unique solvent structure.

such as the amine nitrogen in arginine and lysine (results not shown), are negative due to electronegative water oxygen atoms. Figure 1b shows the solvent charge distribution for alpha-carbons in alanine, asparagine, lysine, and glycine. It demonstrates a slightly smaller solvent excluded volume for the alpha-carbon in glycine, likely due to the absence of a side chain. Thus, glycine's alpha carbon was put in a separate radius group from the rest of the alpha carbons. Similar plots were used to classify all of the radius groups which are shown in Table 2 along with their starting values.

We highlight two distinctions from the Nina et al.[10] optimizations, which also used the solvent charge distribution to approximate starting radii. First, we used nonzero hydrogen radii. This was deemed important because FEP charging free energies deviated 2−13% depending upon hydrogen placement (results not shown). GB models have also shown sensitivity to hydrogen radii.[35] Although the GB and PB models are fundamentally different, both are hopeful methods for DNA dynamics and both will likely need to fine-tuned their hydrogen radii. Second, although the starting radii work well with vdW surface definitions, as used by Nina et al., they clearly underestimated solvation effects with molecular surface definitions. We found that decreasing the starting values by 10% reduced the number of evolutions necessary to converge the radii to their optimal values.

3.2. FEP Simulations. The approximate starting radii were fine-tuned to reproduce explicit solvent FEP charging free energies. These energies, shown in Tables 3 and 4, were used

Table 4. Explicit Solvent Charging Free Energies (WHAM) and Continuum Solvation Energies for the 20 Amino Acid Dipeptides Using AMBER and Optimized Radii[c]

res	WHAM		AMBER	opt[a]	opt[b]
		Nonpolar Groups			
Gly	−13.91	(0.10)	−14.35	−14.43	−13.93
Gly2	−18.45	(0.15)	−18.51	−18.36	−18.11
Ala	−14.01	(0.18)	−13.77	−14.26	−13.82
Ala2	−17.86	(0.02)	−16.97	−16.92	−16.90
Val	−12.48	(0.17)	−11.22	−12.58	−12.45
Val2	−18.79	(0.45)	−16.99	−17.36	−17.73
Leu	−12.58	(0.02)	−11.72	−12.68	−12.40
Leu2	−18.45	(0.06)	−16.77	−17.97	−17.94
Ile	−11.99	(0.06)	−10.87	−12.01	−12.00
Ile2	−19.29	(0.06)	−17.25	−17.52	−17.85
Pro	−14.75	(0.01)	−13.29	−14.76	−14.74
Pro2	−16.84	(0.16)	−14.66	−16.75	−16.98
Phe	−16.42	(0.46)	−14.68	−16.45	−16.43
Phe2	−21.95	(0.01)	−20.16	−21.66	−20.97
Trp	−19.76	(0.04)	−19.50	−19.73	−19.73
Trp2	−22.23	(0.45)	−23.25	−23.14	−21.93
Met	−12.98	(0.02)	−12.76	−13.12	−13.09
Met2	−19.53	(0.23)	−18.40	−18.85	−19.04
		Neutral Polar Groups			
Ser	−17.34	(0.03)	−18.28	−17.33	−17.29
Ser2	−19.60	(0.18)	−21.03	−19.68	−19.82
Thr	−16.25	(0.18)	−17.06	−16.69	−16.52
Thr2	−19.58	(0.30)	−19.68	−19.11	−19.24
Cys	−18.18	(0.36)	−18.69	−18.17	−18.11
Cys2	−17.97	(0.07)	−18.39	−17.96	−17.95
Tyr	−19.78	(0.13)	−20.49	−20.42	−19.77
Tyr2	−25.30	(0.23)	−25.32	−25.08	−24.26
Asn	−20.88	(0.10)	−23.09	−20.89	−20.86
Asn2	−23.38	(0.47)	−24.96	−23.36	−23.12
Gln	−18.40	(0.47)	−20.16	−18.43	−18.68
Gln2	−27.63	(0.24)	−29.29	−26.40	−26.35
His	−23.63	(0.07)	−24.24	−24.77	−24.26
His2	−22.12	(0.16)	−21.13	−22.52	−21.56
		Charged Polar Groups			
Arg	−60.99	(0.44)	−69.19	−63.89	−63.05
Arg2	−77.28	(0.70)	−81.88	−76.10	−74.50
Lys	−65.99	(0.06)	−71.66	−67.44	−68.71
Lys2	−82.65	(1.14)	−85.93	−81.69	−80.33
Asp	−77.86	(0.47)	−72.05	−77.90	−78.02
Asp2	−79.38	(0.18)	−71.06	−77.35	−77.33
Glu	−79.00	(0.51)	−72.09	−79.09	−78.99
Glu2	−76.31	(0.24)	−69.47	−76.30	−76.30

[a] Abrupt smoothed dielectric boundary definitions. [b] Spline smoothed dielectric boundary definitions. [c] All energies are in kcal/mol. Conformation 1 and 2 have phi/psi angles of (180°,180°) and (−40°,-60°), respectively. Standard errors are reported as half the difference between the forward and reverse WHAM energies.

as the target values in the GA optimizations. All simulations were well converged as previously described and as indicated by the reported standard errors. Values for the polyalanine peptides ranged from −31 to −81 kcal/mol depending on their length and conformation, while those of the side chain dipeptides ranged from −12 to −83 kcal/mol. The neutral nonpolar residues had the lowest magnitudes, −12 to −16 kcal/mol; the polar residues were slightly higher, −16 to −28 kcal/mol; and the charged residues were the highest, −61 to −83 kcal/mol. Comparing the charging free energies between dipeptide conformations demonstrates a moderate range of conformational sensitivity. Lysine demonstrated the largest range with a charging free energy of −61 kcal/mol for one conformation and −77 kcal/mol for the other.

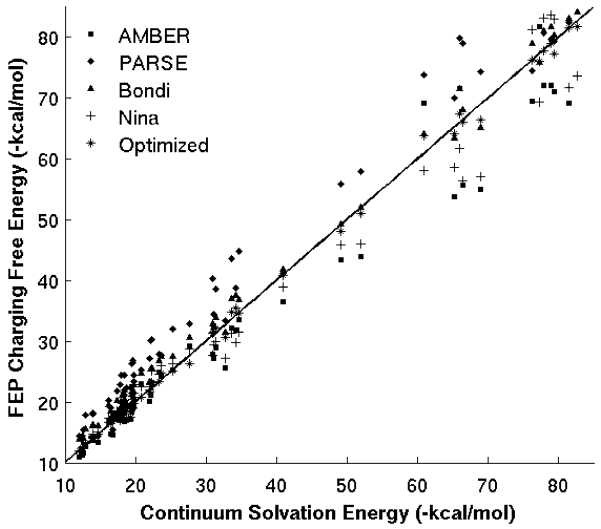

Figure 2. Explicit solvent charging free energies versus continuum solvation free energies calculated with the AMBER charges and different radii.

3.3. GA Optimizations. Results for the GA optimizations were typical for a highly dimensional rough energy landscapes. Specifically, different radius sets with similar fitness values were often encountered. After 6 evolutions the optimal abrupt and smooth boundary radii, as presented in Table 1, had AAEs of 0.54 and 0.64, respectively.

3.4. Continuum Calculations Testing Parameter Sets. To gauge the importance of these optimized radii, we tested several commonly used continuum parameter sets: AMBER (parm99) charges combined with parm99 vdW radii, Bondi radii, and the previously optimized Nina et al. radii in addition to PARSE charges combined with PARSE radii. Figure 2 shows the resulting continuum solvation energies compared to explicit solvent charging energies. The relative performance between parameter sets is more clearly illustrated in Figure 3a where the continuum deviation for each model compound is represented as a single point.

Each parameter set was tested for the appropriate solute dielectric and surface definitions for the comparisons made in Figures 2 and 3a. Failure to use the appropriate solute dielectric and surface definitions results in overestimating or underestimating solvation effects. For example, the PARSE parameters, originally optimized with a solute dielectric of 2, consistently overestimate solvation effects when used with a solute dielectric of 1. Likewise the Nina et al. radii, which were optimized with a vdW surface, underestimate solvation effects with a molecular surface definition. Although changing either the solute dielectric or the surface representation generally increases or decreases solvation energies, it has little effect on the relative solvation energies between different solutes. This is demonstrated in Figure 3b with the Bondi radii where using a vdW surface and solute dielectric of 1 overestimates solvation effects and

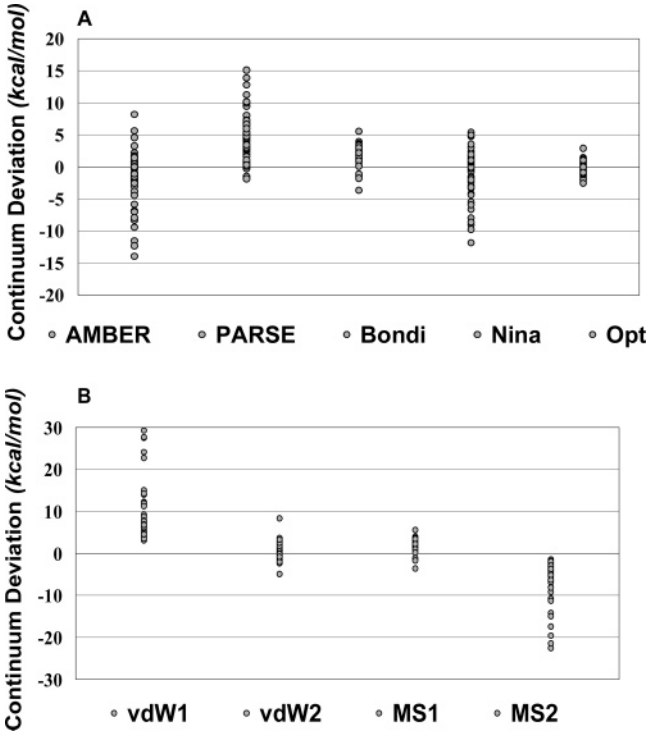

Figure 3. Continuum solvation energy deviations from explicit solvent reference values. Each circle represents one model compound. (A) Five different parameter sets are compared with their optimal surface definitions and solute dielectrics: PARSE charges with PARSE radii, a molecular surface, and solute dielectric of 2; AMBER charges with AMBER radii a molecular surface and solute dielectric of 1; AMBER charges with Bondi radii a molecular surface and solute dielectric of 1; AMBER charges with Nina radii a vdW surface and solute dielectric of 1; and AMBER charges with the newly optimized radii a molecular surface and solute dielectric of 1. (B) The effect of different surface and solute dielectric definitions on solvation energies calculated with the Bondi radii.

a molecular surface with a solute dielectric of 2 underestimates them.

The AMBER parm99 vdW radii were tested because they are frequently chosen for use with parm99 charges despite the fact that they have never been optimized for continuum models. Optimal performance, found with a molecular surface definition and a solute dielectric of 1, yielded an AAE of 3.18 kcal/mol. The Bondi radii[36] were tested because they are the most common choice for the intrinsic radii used in generalized Born models.[37] Using a solute dielectric of 1 and a molecular surface representation they performed the best out of the unoptimized parameters, with an AAE of 2.27 kcal/mol.

The Nina et al. radii were tested to query the transferability of radii between the AMBER and CHARMM force fields. They were not expected to work perfectly with the AMBER charges because continuum charge and radius definitions are strongly interdependent. It has been postulated, however, that the two force fields should have similar solute−solvent interactions, similar charging free energies, and thus similar continuum radii.[38] To test this, FEP simulations were run with the CHARMM force field. They yielded substantially different charging free energies and solvent radial distributions. In continuum calculations using a solute dielectric of 1 and a vdW surface, the Nina et al. radii combined with AMBER charges resulted in an AAE of 2.67 kcal/mol. This is an improvement over parm99 vdW radii but worse than the Bondi radii. It is far worse than their performance with the CHARMM charges when compared to FEP simulations run with the CHARMM force field (results not shown), which had an AAE of 0.69 kcal/mol. Differences in the FEP charging free energies, the solvent radial distribution functions, and the optimal continuum radii for AMBER and CHARMM are likely due to differences in partial charges.

Finally, the PARSE parameters were tested as they have likely been the most respected and frequently used continuum charge and radius definitions since their development in 1994.[8] Their performance relative to the explicit solvent model was worse than expected. Regardless of the surface and solute dielectric definitions, they consistently overestimated solvation effects on the model systems. Optimal performance was found with a molecular surface definition and a solute dielectric of 2 resulting in an AAE of 5.22 kcal/mol.

3.5. Protein Test Cases. The optimized radii clearly perform better on the model compounds for which they were optimized, but this is a biased test. As an unbiased test and to evaluate their transferability to proteins, FEP simulations were run on four short polypeptide chains: Trpcage (1l2y),[39] the C-terminal fragment (41−56) of protein G (2gb1),[40] the C-peptide of ribonuclease A (8rat),[41] and the first helix taken from lysozyme (residues 5−14) with all of the native side chains present. The proteins were prepared and simulated with the same procedures used on the model compounds. Simulations were held to the same requirements for energetic convergence. The resulting charging free energies are shown in Table 5 in addition to continuum deviations with the AMBER, PARSE, Bondi, and newly optimized radii. The surface definitions and solute dielectrics found optimal for the model compounds, as described in the previous section, were used for consistency. The optimized radii perform the best on abrupt boundary surfaces with an AAE of 5.96 kcal/mol. All abrupt boundary radius sets overestimated solvation energies by 20−40% when used with the spline smoothed surfaces (results not shown). The smooth boundary radii, presented in the 9th column, are a significant improvement but still overestimate solvation effects slightly. The source of this systematic overestimation is unknown. It can be compensated for by using a higher solute dielectric, which, as previously mentioned, shifts all solvation energies down but has little effect relative continuum deviations. As shown in the final column, the AAE can be decreased to 4.42 kcal/mol by using a solute dielectric of 1.15 instead of 1.0.

3.6. Spline Smoothed Radii and Atomic Forces. If abrupt boundary radii are used with a spline smoothed dielectric boundary, then solvation forces, similar to solvation energies, tend to be overestimated by 10 to 40% (results not shown). The width of the spline smoothed dielectric transition region, often referred to as the *spline window*, has been shown to effect solvation energies.[11] Thus, the optimal spline smoothed radii will be dependent on the spline window width. Previous optimizations have shown that reasonable quantitative ac-

Table 5. FEP Charging Energies and Continuum Deviations for 4 Protein-like Polypeptides[d]

protein	WHAM-40ps	WHAM-60ps	AMBER	PARSE	Bondi	Nina	opt[a]	opt[b]	opt[c]
trpcage	−309.85 (2.00)	−308.63 (1.89)	−20.47	−27.41	1.10	−2.97	0.51	19.57	−3.32
protein G	−314.37 (0.51)	−314.60 (0.25)	−17.78	14.58	21.55	70.55	15.74	19.72	4.86
ribonucleaseA	−167.77 (1.55)	−165.25 (1.21)	−6.21	38.76	11.69	45.41	3.82	17.37	3.83
lysozyme	257.88 (3.89)	−259.79 (2.53)	4.99	−3.09	1.93	31.16	3.77	5.99	−5.69
AAE			12.36	20.96	9.07	37.52	5.96	15.66	4.43

[a] Abrupt and smoothed dielectric boundary definitions with solute dielectric of 1. [b] Spline smoothed dielectric boundary definitions with solute dielectric of 1. [c] Spline smoothed boundary with solute dielectric of 1.15. [d] All energies in kcal/mol. FEP standard errors are reported as half the difference between the forward and reverse WHAM energies.

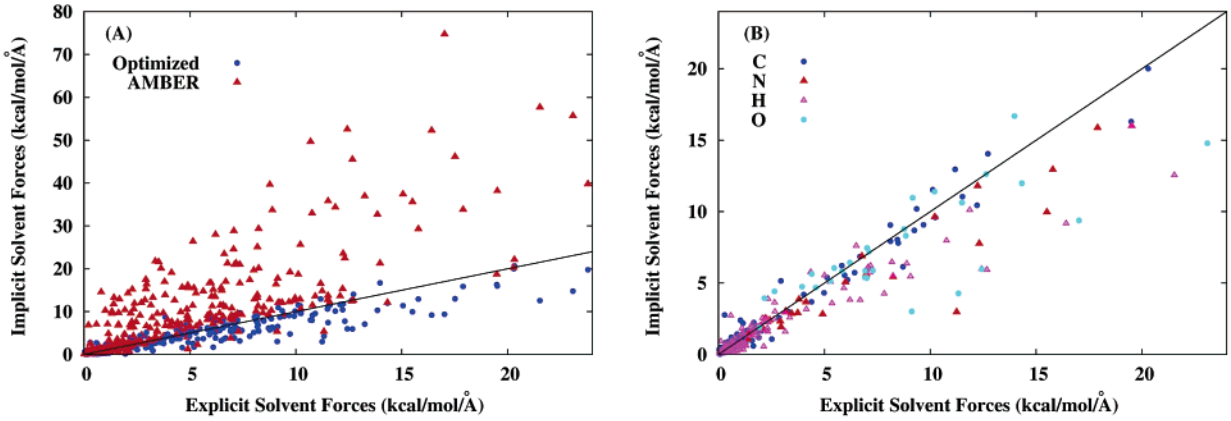

Figure 4. Comparison of explicit solvent and continuum forces. Lines indicate y = x NOT best fit. (A) Continuum forces calculated with AMBER radii in red: correlation coefficient 0.84; slope 1.99; intercept 0.03 kcal mol^{-1} Å^{-1}. Continuum forces calculated with optimized radii in blue: correlation coefficient 0.94; slope 0.80; intercept 0.29 kcal mol^{-1} Å^{-1}. (B) Continuum forces calculated with optimized radii only distinguishing different elements by color.

curacy can be maintained by rescaling abrupt-boundary radii by a single factor that is dependent on the spline window width.[11] This, however, assumes that all radii are equally affected by the spline smoothing. This assumption was tested by optimizing a set of radii for a single spline window of 0.3 Å with the GA. These radii were compared to abrupt boundary radii scaled by a single scaling factor. A simplex optimization showed that the optimal single scaling factor was $R_{spline} = (R_{abrupt} + 0.3)*1.0107$. The independently optimized radii, shown in Table 2, are significantly different from those obtained from the single scaling factor. The AAEs for the independently optimized and single scaled radii were 0.66 and 2.11 kcal/mol, respectively.

Recently, Wagoner et al. suggested comparing continuum and explicit solvent forces instead of molecular solvation energies to test the continuum model.[21] To obtain the explicit solvent electrostatic forces they averaged the total minus nonpolar solvent−solute interactions from a simulation of a 131-residue protein in a fixed conformation. They used the AMBER parm99 vdW radii and demonstrated reasonable correlation between continuum and explicit solvent forces with a correlation coefficient $r = 0.88$. As expected, however, the continuum forces were systematically overestimated by a factor of 2.2, as indicated by the slope from linear regression analysis. To see what effect the optimized radii have on atomic forces, the same procedure was carried out with Trpcage and the C-terminus of G-protein. Explicit solvent forces were reasonably converged after 200 ps of simulation. This was tested by comparing forces from a 500 ps simulation which resulted in a correlation coefficient of 0.9987 and an AAE of 0.10 kcal/mol. Figure 4a shows the improved correlation for atomic forces calculated with the optimized radii, $r = 0.94$, versus the AMBER radii, $r = 0.84$. The AMBER radii still systematically overestimate the continuum forces by a factor of 1.99, while the optimized radii slightly underestimate the forces by a factor of 0.80. Figure 4b shows the optimized atomic forces according to atom type and demonstrates that forces on oxygen atoms are most often underestimated. Comparing specific atomic forces from continuum and explicit solvent models may prove useful in future optimization efforts.

4. Conclusion

Computational biochemistry is a field with multiple levels of models and theories whose computational requirements scale proportionally with accuracy. The success of the lower level, less accurate, more efficient models relies upon their connection to the higher level, more accurate, less efficient models, while the latter are complimented by the former to characterize complex biological systems. As methodology advances, opportunities for improved synergy between levels of theory and models are created. As microscopic, explicit solvent simulations become more accurate and the numerical methods used in continuum models improve, there is an increasing opportunity and need to benchmark continuum models on microscopic simulations.

The main results of this paper are two sets of optimized continuum radii, presented in Table 2, for PB calculations with the AMBER force field using either abrupt or spline smoothed dielectric boundary definitions. Thirty-one radius

groups were defined based on similar chemical environments. Both sets of optimized radii improve quantitative agreement with microscopic simulations when compared to AMBER parm99 vdW, Bondi, and Nina et al. radii as well as the PARSE parameters. The Bondi radii combined with a protein dielectric of 1 and a molecular surface definition were the best alternative choice for AMBER charges. The PARSE parameters, shown to work optimally with a solute dielectric of 2 and a molecular surface definition, were less accurate. The abrupt boundary radii work quite well for protein systems. The smooth boundary radii offer a significant improvement over abrupt boundary radii used on spline smoothed surfaces but still overestimated the solvation energies of four protein-like polypeptides. This overestimation can be compensated for by using a solute dielectric of 1.15 instead of 1.0. The smooth boundary radii greatly improve agreement between continuum and explicit solvent atomic forces.

Future efforts will focus on identifying the source of deviations of the smoothed boundary radii on protein systems as well as small molecule parametrization and coupling these parameters with nonpolar solvation effects. If one hopes to use these radii on systems involving small molecule ligands, it will be important to characterize the appropriate radii for certain small molecules as well. It will also be of great use to test the complementarity of these electrostatic parameters with various nonpolar solvation models by comparing total solvation energies to either explicit solvent simulations or experimental solvation energies. It is expected that the nonpolar and polar contributions to continuum solvation models will be coupled and that the greatest degree of accuracy will require them to be treated as such.

Acknowledgment. The authors thank Dr. Nathan Baker, Dr. Benoit Roux, Dr. Dave Case, and Dr. Michael Feig for valuable discussion and guidance; Dr. Thomas Cheatham III and Dr. Jeffrey Klauda for the AMBER parm99 force field converted to format; and Dr. Robert Konecny and the Center for Theoretical Biological Physics for computing resources. J.M.J.S is supported by the Center for Theoretical Biological Physics. Additional funding for this work comes from National Science Foundation, National Institutes of Health, National Biomedical Computational Resource, Accelrys, Inc., and the Howard Hughes Medical Institute.

Supporting Information Available: Model conformations. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Jackson, J. D. *Classical Electrodynamics*; John Wiley and Sons: 1962; Ch. 1, pp 12−14.

(2) Warwicker, J.; Watson, H. C. *J. Mol. Biol.* **1982**, *157*, 671−679.

(3) Gouy, M. *J. Phys.* **1910**, *9*, 457−468.

(4) Roux, B.; Simonson, T. *Biophys. Chem.* **1999**, *78*, 1−20.

(5) Hirata, F.; Redfern, P.; Levy, R. M. *Int. J. Quantum Chem.* **1988**, 179−190.

(6) Roux, B.; Yu, H. A.; Karplus, M. *J. Phys. Chem.* **1990**, *94*, 4683−4688.

(7) Jayaram, B.; Fine, R.; Sharp, K.; Honig, B. *J. Phys. Chem.* **1989**, *93*, 4320−4327.

(8) Sitkoff, D.; Sharp, K.; Honig, B. *J. Phys. Chem.* **1994**, *98*, 1978−1988.

(9) Bordner, A. J.; Cavasotto, C. N.; Anagyan, R. A. *J. Phys. Chem.* **2002**, *106*, 11009−11015.

(10) Nina, M.; Beglov, D.; Roux, B. *J. Phys. Chem. B* **1997**, *101*, 5239−5248.

(11) Nina, M.; Im, W.; Roux, B. *Biophys. Chem.* **1999**, *78*, 89−96.

(12) Gallicchio, E.; Kubo, M. M.; Levy, R. M. *J. Phys. Chem.* **2000**, *104*, 6271−6285.

(13) Shirts, M. R.; Pitera, J. W.; Swope, W. C.; Pande, V. S. *J. Chem. Phys.* **2003**, *119*, 5740−5761.

(14) Deng, Y. Q.; Roux, B. *J. Phys. Chem. B* **2004**, *108*, 16567−16576.

(15) Massova, I.; Kollman, P. A. *Perspect. Drug Discovery Des.* **2000**, *18*, 113−135.

(16) Swanson, J. M. J.; Henchman, R. H.; McCammon, J. A. *Biophys. J.* **2004**, *86*, 67−74.

(17) Chipman, D. M. *J. Phys. Chem. A* **2002**, *106*, 7413−7422.

(18) Beroza, P.; Case, D. A. *Energet. Biol. Macromol., B* **1998**, *295*, 170−189.

(19) Lu, Q.; Luo, R. *J. Chem. Phys.* **2003**, *119*, 11035−11047.

(20) Mongan, J.; Case, D. A. *Curr. Opin. Struct. Biol.* **2004**.

(21) Wagoner, J. Baker, N. A. *J. Comput. Chem.* **2004**.

(22) Grant, J. A.; Pickup, B. T.; Nicholls, A. *J. Comput. Chem.* **2001**, *22*, 608−640.

(23) Im, W.; Beglov, D.; Roux, B. *Comput. Phys. Commun.* **1998**, *111*, 59−75.

(24) Baker, N. A.; Sept, D.; Joseph, S.; Holst, M. J.; McCammon, J. A. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 10037−10041.

(25) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187−217.

(26) Luo, R.; David, L.; Gilson, M. K. *J. Comput. Chem.* **2002**, *23*, 1244−1253.

(27) Michael, F.; Karanicolas, J.; Brooks, C. L., III. *MMTSB Toolset*; MMTSB NIH Research Resource: The Scripps Research Institute, 2001.

(28) Dunbrack, R. L.; Cohen, F. E. *Protein Sci.* **1997**, *6*, 1661−1681.

(29) Laskowski, R. A.; Hutchinson, E. G.; Michie, A. D.; Wallace, A. C.; Jones, M. L.; Thornton, J. M. *Trends Biochem. Sci.* **1997**, *22*, 488−490.

(30) Beglov, D.; Roux, B. *J. Chem. Phys.* **1994**, *100*, 9050−9063.

(31) Nina, M.; Simonson, T. *J. Phys. Chem. B* **2002**, *106*, 3696−3705.

(32) Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M. *J. Comput. Chem.* **1992**, *13*, 1011−1021.

(33) Dong, F.; Vijayakumar, M.; Zhou, H. X. *Biophys. J.* **2003**, *85*, 49−60.

(34) Baker, J. E. Reducing bias and inefficiency in the selection algorithm. In *Proceedings of the Second International Conference on Genetic Algorithms on Genetic algorithms and their application*; Lawrence Erlbaum Associates, Inc.: Mahwah, 1987; pp 14−21.

(35) Tsui, V.; Case, D. A. *J. Am. Chem. Soc.* **2000**, *122*, 2489−2498.

(36) Bondi, A. *J. Chem. Phys.* **1964**, *64*, 441.

(37) Bashford, D.; Case, D. *Annu. Rev. Phys. Chem.* **2000**, *51*, 129−152.

(38) Roux, B.; Case, D. personal communication.

(39) Neidigh, J. W.; Fesinmeyer, R. M.; Andersen, N. H. *Nature Struct. Biol.* **2002**, *9*, 425−430.

(40) Gronenborn, A. M.; Filpula, D. R.; Essig, N. Z.; Achari, A.; Whitlow, M.; Wingfield, P. T.; Clore, G. M. *Science* **1991**, *253*, 657−661.

(41) Tilton, R. F.; Dewan, J. C.; Petsko, G. A. *Biochemistry* **1992**, *31*, 2469−2481.

CT049834O