THE JOURNAL OF PHYSICAL CHEMISTRY B

ARTICLE

pubs.acs.org/JPCB

A Refined, Efficient Mean Solvation Force Model that Includes the Interior Volume Contribution

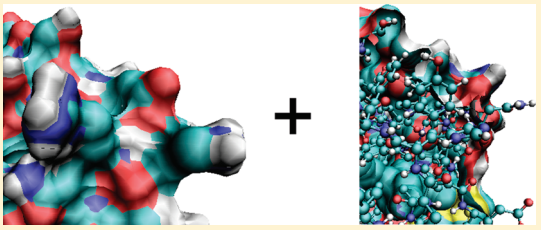
Jane R. Allison,[†] Katharina Boguslawski,[†] Franca Fraternali,[‡] and Wilfred F. van Gunsteren*[,†]

[†]Laboratory of Physical Chemistry, Swiss Federal Institute of Technology ETH, 8093 Zürich, Switzerland

[‡]Randall Division of Cell and Molecular Biophysics, King's College London, London SE1 1UL, United Kingdom

Supporting Information

ABSTRACT: A refined implicit aqueous solvation model is proposed for the simulation of biomolecules without the explicit inclusion of the solvent degrees of freedom. The mean force due to solvation is approximated by the derivative of a simple analytic function of the solvent accessible surface area combined with two atomic solvation parameters, as described previously, with the addition of a novel term to account for the interaction of the interior atoms of the solute with the solvent. The extended model is parametrized by comparing the structural properties and energies computed from simulations of six test proteins of varying sizes and shapes using the new solvation energy term with the corresponding values obtained from simulations in vacuum, using the original implicit solvent model and in explicit water, and from the X-ray or NMR model structures. The mean solvation model proposed here improves the structural properties relative to vacuum simulations and relative to the simpler model that neglects the volume contribution, while remaining significantly more efficient than simulations in explicit water.



Molecular simulation is widely used to investigate the structure and dynamics of biological macromolecules. Typically, for single macromolecules, an atomic-level representation is used. Inclusion of the solvent then results in a huge number of degrees of freedom, greatly increasing the cost of the calculations. Because of this, the earliest simulations of biomolecules took place in vacuum. Complete neglect of the solvent effects, however, greatly limits the accuracy of the simulation.[1] Computational power has now increased sufficiently that in many cases it is feasible to do all-atom molecular dynamics (MD) simulations in explicit solvent. For large-scale or computationally expensive applications such as drug screening, protein structure prediction, protein folding, computational engineering of peptides and proteins, and aggregation studies, however, a compromise between the speed of simulations done in vacuum and the accuracy of explicitly representing the solvent is required. Implicit solvent models, in which the effect of the solvent is modeled as a mean solvation term in the potential energy function, provide a means of including solvent effects at minimal cost.[2−5]

The form and parameters of the solvation term may be derived in a range of different ways and at varying levels of accuracy. A widely used and simple class of implicit solvent models are the so-called SASA (solvent accessible surface area) models, in which the local solute−solvent interactions are assumed to be proportional to the SASA of the solute atoms.[6−20] Each atom type is assigned a solvation parameter that reflects its hydrophobicity or hydrophilicity. Electrostatic interactions may be accounted for by either neutralizing the charged moieties,[16] or supplementing the SASA contribution to the energy function with a screened Coulomb energy term, as in the CASA (Coulomb/accessible surface area) models.[21] These models have been shown to have an accuracy comparable to more sophisticated theoretical models.[21]

The SASA term should account for the free energy cost of forming a hard-sphere cavity in water. It has been shown, however, that the cavity creation work depends on both the solvent accessible volume and the SASA, with a crossover to SASA dominance at large solute sizes.[22−27] Another factor missing from existing SASA-based implicit solvent models is the inclusion of the favorable van der Waals interactions between the interior atoms of the solute and the solvent.[28−36]

Previously, a SASA-based implicit solvent model[16] for proteins in water was incorporated into the GROMOS biomolecular simulation package.[37] The same model, with a virtually identical parametrization, was later also used in conjunction with the CHARMM force field.[18] In the spirit of simplification, a functional form with a minimal number of parameters was chosen: electrostatics were accounted for by neutralizing the charged moieties and just two atomic solvation parameters, for hydrophilic and hydrophobic atom types, were used.

Here, the GROMOS SASA-based implicit solvent model is extended to take into account interactions between buried atoms and the solvent as well. A term proportional to the internal volume, defined as the sum of the volumes of all nonsolvent-exposed atoms, is added. Just one additional parameter is required. Moreover, because the atomic surface areas are used to define which atoms contribute to the internal volume, the computational efficiency of the original model is retained. The new SASA/VOL implicit solvent

Received: February 21, 2011
Published: March 24, 2011

model is parametrized by comparison of structural features computed as averages over trajectories generated by simulation of six proteins of different size and shape (Table 1) in vacuum, using the original SASA model, in explicit water and using the SASA/VOL solvation term, as well as from the X-ray or NMR model structures. The improved solvation force model is better able to reproduce these key features and generates stable protein structures, while remaining computationally efficient.

THEORY

Functional Form of the SASA/VOL Mean Solvation Energy Term. The solvation free enthalpy of a solute G_{solv} may be partitioned into a solvent−solvent cavity term G_{cav}, a solute−solvent van der Waals term G_{vdW}, and a solute−solvent electrostatic term G_{elec}:

$$G_{solv} = G_{cav} + G_{vdW} + G_{elec} \quad (1)$$

Here, as in the original SASA model built upon the GROMOS force field,[16] rather than calculate G_{elec} explicitly, the vacuum parametrization for the intrasolute interaction function is used: the partial charges of groups of atoms with nonzero total charge are changed such that these groups have zero charge, but their hydrogen-bonding capacity is maintained.[38] G_{cav} and G_{vdW} are combined into a mean solvation term, such that the total interaction function for a protein solute containing N atoms with Cartesian coordinates $\mathbf{r}^N = (\mathbf{r}_1, \mathbf{r}_2, ..., \mathbf{r}_N)$ consists of two parts:

$$V(\mathbf{r}^N) = V_{phys}(\mathbf{r}^N) + V_{solv}(\mathbf{r}^N) \quad (2)$$

where $V_{phys}(\mathbf{r}^N)$ is determined by a standard GROMOS force field for vacuum simulations.[38] The mean solvation term developed here is further split:

$$V_{solv}(\mathbf{r}^N) = V_{solv}^{SASA}(\mathbf{r}^N) + V_{solv}^{VOL}(\mathbf{r}^N) \quad (3)$$

Each term in eq 3 is calculated as a sum of atomic contributions. The first term is that of the original GROMOS SASA model:[16]

$$V_{solv}^{SASA}(\mathbf{r}^N) = \sum_{i=1}^{N} \sigma_i^{SASA} A_i(\mathbf{r}^N) \quad (4)$$

There are just two types of atom-specific solvation parameters σ_i, for hydrophilic ($\sigma_{O,N}^{SASA}$) and hydrophobic (σ_C^{SASA}) atoms (Table 2). The SASA of atom i, A_i, is defined using the approximate analytical expression of Hasel et al.:[39]

$$A_i(\mathbf{r}^N) = S_i \prod_{\substack{j=1 \\ j\neq i}}^{N} \left[1 - p_i p_{ij} \frac{b_{ij}(r_{ij})}{S_i}\right] \quad (5)$$

The total surface area of an isolated atom i with radius R_i accessible to a solvent probe atom with radius R_{solv} is given by

$$S_i = 4\pi[R_i + R_{solv}]^2 \quad (6)$$

Table 1. Abbreviated Names, PDB Entry Codes, and Some Properties of the Proteins Used To Parametrize the SASA/VOL Implicit Solvation Model

name	PDB code	source	N_{res}[a]	N_{atoms}[b]	N_{hbonds}[c]	SS[d]	size/shape
trp	1L2Y	NMR	20	198	184	α-helix	disk
drk	2A36	NMR	59	610	506	β-sheet	distorted sphere
ubq	1UBQ	X-ray	76	760	788	α/β	sphere
if3c	1TIG	X-ray	88	901	888	α/β	elongated
lys	1AKI	X-ray	129	1331	1368	α/β	deformed sphere
talin	2JSW	NMR	189	1699	1580	α-helix	elongated

[a] Number of residues in the protein. [b] Number of atoms in the GROMOS SASA/VOL representation of the protein. [c] Number of hydrogen bonds in the energy-mimized X-ray or NMR model structure. [d] Predominant secondary structure class.

Table 2. GROMOS Atom Types and the Corresponding SASA/VOL Mean Solvation Parameters

atom type[a]	R_i[b]	p_i[b]	σ_i^{SASA}[c]	description
OA; OW; OE	0.152	1.080	−40	hydroxyl oxygen (OH); water oxygen; ester oxygen
O	0.150	0.926	−40	carbonyl (C═O)
OM	0.170	0.922	−40	carboxyl (C—O^-)
NT; NL	0.160	1.215	−40	terminal nitrogen (NH_2); (NH_3)
N; NR; NZ; NE	0.155	1.028	−40	peptide nitrogen (N or NH); aromatic N; Arg NH (NH_2); Arg NE (NH)
C; CH0	0.172	1.554	5	bare carbon (peptide, C═O, C—N); bare sp^3 carbon with four bound non-hydrogen atoms
CH1	0.180	1.276	5	aliphatic CH group
CH2; CH2r	0.190	1.045	5	aliphatic CH_2 group; CH_2 in a ring
CH3	0.200	0.880	5	aliphatic CH_3 group
CR1	0.180	1.073	5	aromatic CH group
H	0.110	1.128	0	hydrogen
S	0.180	1.121	0	sulfur

neighbor type	p_{ij}[b]
first (one bond)	0.8875
second (two bonds)	0.3316
third and higher (three or more bonds)	0.3316

[a] As in the GROMOS 45A4[41] and 53A6[42] force fields and their corresponding vacuum formulations 45B4 and 53B6, respectively. Only the atom types relevant for proteins are listed. [b] Used in the approximate expression for the SASA (eqs 5−7); taken from Hasel et al.[39] [c] Values chosen for use with $\sigma^{VOL} = -100$ kJ·mol^{-1}·nm^{-3} in the SASA/VOL mean solvation model according to the tests described in this work. In the previously published SASA-only mean solvation model,[16] $\sigma_{O,N}^{SASA} = -25$ kJ·mol^{-1}·nm^{-2} and $\sigma_C^{SASA} = 5$ kJ·mol^{-1}·nm^{-2}.

dx.doi.org/10.1021/jp2017117 |J. Phys. Chem. B 2011, 115, 4547–4557

Table 3. Calculated Properties for the trp (20 residues) NMR Model Structure and for Simulations in Explicit Water, in Vacuum, and with the SASA and SASA/VOL Implicit Solvation Models

					SASA[c]			energies[d]				
references	rmsd[a]	rmsf[a]	R_{gyr}[a]	%H-bonds[b]	phob	phil	total	vdw(u:u)	ele(u:u)	ele(u:v)	vdw(u:v)	totpot
NMR			0.74	100	7.9	5.5	15.4	−493	−627			−787
H_2O (20 ns)	0.34	0.17	0.75	92	8.6	5.8	16.6	−379	−851	−2791	−229	−114206
SASA/VOL (10 ns)	0.32	0.15	0.74	97	8.2	5.4	15.4	−430	−1308	−173	−86	−1084
SASA (10 ns)	0.34	0.20	0.76	94	8.4	4.9	15.1	−417	−1322	−81		−1081
vacuum (100 ns)	0.39	0.20	0.70	99	8.2	4.0	13.7	−445	−1390			−1149

parametrization[e]							SASA[c]			energies[d]				
$\sigma_{O,N}^{SASA}$[f]	σ_{C}^{SASA}[f]	σ^{VOL}[f]	rmsd[a]	rmsf[a]	R_{gyr}[a]	% H-bonds[b]	phob	phil	total	vdw(u:u)	ele(u:u)	sasa	vol	totpot
−30	5	−80	0.17	0.06	0.70	102	7.8	4.8	14.3	−431	−1322	−106	−70	−1105
−30	5	−100	0.16	0.07	0.70	102	8.0	5.0	14.8	−436	−1295	−110	−87	−1084
−30	5	−120	0.21	0.07	0.68	103	8.1	5.0	14.8	−439	−1304	−111	−105	−1083
−40	5	−80	0.25	0.07	0.73	96	8.6	5.9	16.6	−425	−1231	−194	−63	−1019
−40	5	−100	0.15	0.08	0.72	102	8.1	5.1	15.0	−430	−1293	−165	−87	−1078
−40	5	−120	0.13	0.07	0.72	101	8.2	5.3	15.3	−416	−1290	−172	−104	−1071
−50	5	−80	0.14	0.06	0.73	99	8.5	5.7	16.2	−408	−1273	−243	−65	−1051
−50	5	−100	0.14	0.07	0.74	99	8.4	5.7	16.1	−412	−1272	−241	−81	−1048
−50	5	−120	0.15	0.07	0.73	99	8.3	5.7	16.0	−417	−1258	−242	−99	−1036

[a] C_α atom-positional root-mean-square deviations (rmsd) and atom-positional root-mean-square fluctuations (rmsf) in nm. Each structure in the trajectory was first superimposed on the X-ray or NMR model structure energy-minimized in the corresponding force field by minimizing the rmsd of the C_α atoms. All atoms were included in the calculation of the radius of gyration (R_{gyr}), which, for the simulations, is averaged over the trajectory. [b] The percentage of the number of hydrogen bonds in the X-ray or NMR model structure, averaged over the trajectory in the case of the simulations. [c] Hydrophobic and hydrophilic solvent accessible surface areas (SASA) in nm^2, calculated using the same formula as is used in the SASA and SASA/VOL implicit solvation models and, for the simulations, averaged over the trajectory. Atoms were considered to be hydrophobic or hydrophilic if they would be assigned σ_C^{SASA} or $\sigma_{O,N}^{SASA}$, respectively. Hydrogen atoms were not included in the calculation. [d] Energies are in $kJ\cdot mol^{-1}$ and are trajectory averages for the simulations. The SASA and VOL energies were not calculated for the references because they depend on the value of σ^{VOL}. "u:u" refers to the intrasolute energy terms, "u:v" to the solute:solvent terms, and "totpot" to the total potential energy. For the SASA reference simulation and the SASA/VOL simulation using the optimized parameters, the SASA and VOL energies rather than the ele(u:v) and vdw(u:v) energies are given. [e] For the parametrization simulations, the quantities reported are averages over the last 300 ps of the 500 ps simulation. For the reference simulations, the average is over the time shown in brackets. The simulation using the optimized SASA/VOL parameters started from the end of the parametrization simulation. [f] The $\sigma_{O,N}^{SASA}$ and σ_C^{SASA} values are in $kJ\cdot mol^{-1}\cdot nm^{-2}$, and the σ^{VOL} values are in $kJ\cdot mol^{-1}\cdot nm^{-3}$. In the SASA simulation, $\sigma_{O,N}^{SASA} = -25$ and $\sigma_C^{SASA} = 5$ as in the original formulation of the SASA implicit solvent model,[16] and in the SASA/VOL simulation, the optimized parameters $\sigma_{O,N}^{SASA} = -40$, $\sigma_C^{SASA} = 5$ and $\sigma^{VOL} = -100$ were used.

and the overlap reduction factor[40] b_{ij} for atoms i and j at a distance $r_{ij} = ((\mathbf{r}_i - \mathbf{r}_j)^2)^{1/2}$ is given by

$$b_{ij}(r_{ij}) = \begin{cases} 0 & \text{if} \quad r_{ij} \geq R_i + R_j + 2R_{solv} \\ \pi[R_i + R_{solv}][R_i + R_j + 2R_{solv} - r_{ij}] \times \left[1 + \dfrac{R_j - R_i}{r_{ij}}\right] & \text{if} \quad 0 < r_{ij} < R_i + R_j + 2R_{solv} \end{cases} \tag{7}$$

The parameters p_i and p_{ij} were optimized by Hasel et al.[39] for $R_{solv} = 0.14$ nm. Their mapping onto GROMOS atom types, along with the corresponding R_i values, are given in Table 2.

The SASA values A_i obtained in eq 5 are used in a switching function $g(A)$ to select the atoms of the solute that have $A_i = 0$ to contribute to the novel $V_{solv}^{VOL}(\mathbf{r}^N)$ term of eq 3. This term accounts for the contribution of the interior solute atoms to the solvation free energy:[31]

$$V_{solv}^{VOL}(\mathbf{r}^N) = \sum_{i=1}^{N} \sigma_i^{VOL} g(A_i(\mathbf{r}^N); A_i^{s_1}, A_i^{s_2}) \frac{4}{3}\pi R_i^3 \tag{8}$$

The parameters $A_i^{s_1}$ and $A_i^{s_2}$ switch g from 1 to 0 for $A_i^{s_1} \leq A_i \leq A_i^{s_2}$, so that only atoms with $A_i \leq A_i^{s_2}$ contribute to the buried volume. The conditions for the choice of switching function are (with $0 \leq A_i^{s_1} \leq A_i^{s_2}$)

$$\begin{aligned} g(A_i^{s_1}) &= 1 \quad \frac{dg}{dA_i}(A_i^{s_1}) = 0 \\ g(A_i^{s_2}) &= 0 \quad \frac{dg}{dA_i}(A_i^{s_2}) = 0 \end{aligned} \tag{9}$$

The following switching function was chosen:

$$g(A_i; A_i^{s_1}, A_i^{s_2}) = \begin{cases} 1 & 0 \leq A_i \leq A_i^{s_1} \\ \dfrac{2[A_i - A_i^{s_1}]^3}{[A_i^{s_2} - A_i^{s_1}]^3} - \dfrac{3[A_i - A_i^{s_1}]^2}{[A_i^{s_2} - A_i^{s_1}]^2} + 1 & A_i^{s_1} \leq A_i \leq A_i^{s_2} \\ 0 & A_i^{s_2} \leq A_i \end{cases} \tag{10}$$

This differentiable approximation of a step function gives the value 1 if an atom has a very small or zero A_i, i.e., the atom is in the middle of the solute, and a value of zero if an atom is at the surface, i.e., A_i is large and thus the atom should not contribute to the buried volume term.

The choices of the parameters that can be freely adjusted in the SASA/VOL model, namely, $A_i^{s_1}$, $A_i^{s_2}$, σ_i^{SASA}, and σ_i^{VOL}, are discussed in Results and Discussion. The spatial derivatives of the equations constituting the mean solvation energy term, which are required to compute the forces, are given in the Supporting Information.

dx.doi.org/10.1021/jp2017117 |J. Phys. Chem. B 2011, 115, 4547–4557

Table 4. Calculated Properties for the drk Domain (59 residues) NMR Model Structure and for Simulations in Explicit Water, in Vacuum, and with the SASA and SASA/VOL Implicit Solvation Models

					SASA[c]			energies[d]				
references	rmsd[a]	rmsf[a]	R_{gyr}[a]	% H-bonds[b]	phob	phil	total	vdw(u:u)	ele(u:u)	ele(u:v)	vdw(u:v)	totpot
NMR			1.06	100	21.2	20.1	47.3	−1787	−1521			−2546
H_2O(20 ns)	0.31	0.13	1.11	101	21.9	20.6	48.6	−1560	−3052	−11150	−243	−222085
SASA/VOL (10 ns)	0.25	0.10	1.04	102	20.1	19.3	44.1	−1862	−2415	−672	−289	−2590
SASA (10 ns)	0.34	0.13	1.06	99	21.6	18.5	45.2	−1851	−2400	−355		−2561
vacuum (100 ns)	0.36	0.13	1.02	104	20.1	13.9	38.1	−1979	−2551			−2793

parametrization[e]							SASA[c]			energies[d]				
$\sigma_{O,N}^{SASA}$[f]	σ_{C}^{SASA}[f]	σ^{VOL}[f]	rmsd[a]	rmsf[a]	R_{gyr}[a]	% H-bonds[b]	phob	phil	total	vdw(u:u)	ele(u:u)	sasa	vol	totpot
−30	5	−80	0.21	0.06	1.06	99	20.9	19.1	45.1	−1840	−2340	−470	−229	−2501
−30	5	−100	0.22	0.06	1.04	104	20.2	18.5	43.1	−1863	−2390	−453	−302	−2567
−30	5	−120	0.17	0.06	1.05	103	20.3	18.4	43.7	−1856	−2351	−450	−357	−2528
−40	5	−80	0.20	0.06	1.06	102	21.0	19.7	45.7	−1847	−2332	−685	−225	−2467
−40	5	−100	0.20	0.07	1.06	100	21.0	20.3	46.7	−1820	−2292	−706	−275	−2441
−40	5	−120	0.17	0.05	1.05	101	20.3	19.3	44.6	−1875	−2306	−670	−355	−2491
−50	5	−80	0.23	0.07	1.08	100	22.0	21.4	48.8	−1792	−2267	−959	−208	−2367
−50	5	−100	0.30	0.07	1.07	104	21.3	20.9	47.3	−1780	−2343	−939	−270	−2427
−50	5	−120	0.25	0.06	1.07	100	20.9	20.7	47.2	−1791	−2305	−932	−324	−2370

[a−f] See footnote to Table 3 for explanations of the quantities described here.

Table 5. Calculated Properties for the ubq (76 residues) X-ray Structure and for Simulations in Explicit Water, in Vacuum, and with the SASA and SASA/VOL Implicit Solvation Models

					SASA[c]			energies[d]				
references	rmsd[a]	rmsf[a]	R_{gyr}[a]	% H-bonds[b]	phob	phil	total	vdw(u:u)	ele(u:u)	ele(u:v)	vdw(u:v)	totpot
X-ray			1.19	100	28.0	22.3	56.5	−2477	−2734			−4417
H_2O (10 ns)	0.26	0.12	1.21	80	30.0	23.2	60.6	−1960	−4148	−11286	−357	−321212
SASA/VOL (10 ns)	0.35	0.15	1.17	77	27.1	21.9	55.2	−2328	−3245	−741	−319	−3371
SASA (10 ns)	0.29	0.09	1.14	78	28.7	20.3	54.7	−2382	−3273	−363		−3477
vacuum (100 ns)	0.47	0.13	1.13	84	27.8	15.6	48.0	−2522	−3490			−3779

parametrization[e]							SASA[c]			energies[d]				
$\sigma_{O,N}^{SASA}$[f]	σ_{C}^{SASA}[f]	σ^{VOL}[f]	rmsd[a]	rmsf[a]	R_{gyr}[a]	% H-bonds[b]	phob	phil	total	vdw(u:u)	ele(u:u)	sasa	vol	totpot
−30	5	−80	0.21	0.06	1.15	80	27.7	19.7	53.3	−2385	−3301	−451	−264	−3474
−30	5	−100	0.21	0.07	1.15	79	27.8	19.9	53.3	−2356	−3304	−459	−335	−3465
−30	5	−120	0.26	0.05	1.12	81	26.7	19.2	51.2	−2430	−3304	−441	−431	−3534
−40	5	−80	0.19	0.07	1.18	78	28.3	21.8	56.5	−2294	−3197	−732	−251	−3275
−40	5	−100	0.23	0.06	1.17	77	27.9	21.9	56.4	−2335	−3170	−737	−320	−3321
−40	5	−120	0.22	0.08	1.17	80	27.9	22.0	56.0	−2311	−3176	−741	−388	−3292
−50	5	−80	0.18	0.06	1.20	76	29.5	23.4	59.8	−2269	−3164	−1024	−231	−3247
−50	5	−100	0.22	0.07	1.18	77	29.2	23.5	59.5	−2258	−3152	−1030	−298	−3208
−50	5	−120	0.21	0.07	1.15	76	28.7	22.9	57.9	−2287	−3178	−1002	−372	−3285

[a−f] See footnote to Table 3 for explanations of the quantities described here.

METHODS

The parametrization protocol is based on the comparison of structural properties calculated from the X-ray or NMR model structures and from simulations in vacuum, using the original SASA model,[16] in explicit water and using the SASA/VOL implicit solvation model of six proteins of different sizes and shapes (Table 1). All simulations were carried out using the GROMOS biomolecular simulation package.[37] A full description is given in the Supporting Information.

The total simulation time for the vacuum simulations was 100 ns, except for talin (2JSW), for which 40 ns were obtained in view of the large size of this protein (169 residues/1699 atoms).

dx.doi.org/10.1021/jp2017117 |J. Phys. Chem. B 2011, 115, 4547–4557

Table 6. Calculated Properties for the if3c (88 residues) X-ray Structure and for Simulations in Explicit Water, in Vacuum, and with the SASA and SASA/VOL Implicit Solvation Models

					SASA[c]			energies[d]				
references	rmsd[a]	rmsf[a]	R_{gyr}[a]	% H-bonds[b]	phob	phil	total	vdw(u:u)	ele(u:u)	ele(u:v)	vdw(u:v)	totpot
X-ray			1.29	100	31.6	22.3	60.9	−2973	−3459			−5616
H_2O (10 ns)	0.18	0.11	1.34	86	35.7	26.0	71.5	−2309	−4503	−14051	−443	−336545
SASA/VOL (10 ns)	0.22	0.08	1.27	86	32.8	24.8	65.7	−2779	−4037	−828	−395	−4277
SASA (10 ns)	0.30	0.14	1.28	84	34.0	24.6	66.6	−2766	−3989	−445		−4252
vacuum (100 ns)	0.37	0.18	1.22	88	31.3	18.2	56.0	−2982	−4134			−4538

parametrization[e]							SASA[c]			energies[d]				
$\sigma_{O,N}^{SASAf}$	σ_{C}^{SASAf}	σ^{VOLf}	rmsd[a]	rmsf[a]	R_{gyr}[a]	% H-bonds[b]	phob	phil	total	vdw(u:u)	ele(u:u)	sasa	vol	totpot
−30	5	−80	0.21	0.07	1.27	86	32.6	23.8	63.8	−2798	−3954	−549	−316	−4255
−30	5	−100	0.21	0.07	1.28	86	32.5	23.8	64.1	−2794	−3956	−552	−400	−4233
−30	5	−120	0.19	0.07	1.27	88	32.3	23.4	63.0	−2789	−3988	−540	−495	−4277
−40	5	−80	0.20	0.08	1.29	84	33.9	25.8	68.1	−2733	−3863	−860	−302	−4087
−40	5	−100	0.21	0.07	1.28	86	33.6	25.1	66.9	−2729	−3929	−834	−384	−4158
−40	5	−120	0.21	0.07	1.27	86	32.7	24.7	64.8	−2770	−3957	−822	−485	−4203
−50	5	−80	0.20	0.08	1.29	84	34.5	26.4	69.5	−2730	−3884	−1146	−299	−4096
−50	5	−100	0.21	0.07	1.29	84	34.4	26.7	69.8	−2711	−3870	−1162	−377	−4081
−50	5	−120	0.22	0.07	1.28	85	33.9	26.7	69.3	−2720	−3900	−1166	−451	−4101

[a−f] See footnote to Table 3 for explanations of the quantities described here.

Table 7. Calculated Properties for the lys (129 residues) X-ray Structure and for Simulations in Explicit Water, in Vacuum, and with the SASA and SASA/VOL Implicit Solvation Models

					SASA[c]			energies[d]				
references	rmsd[a]	rmsf[a]	R_{gyr}[a]	% H-bonds[b]	phob	phil	total	vdw(u:u)	ele(u:u)	ele(u:v)	vdw(u:v)	totpot
X-ray			1.42	100	32.0	27.7	70.9	−4375	−6084			−9127
H_2O (10 ns)	0.23	0.15	1.41	91	37.2	30.8	81.3	−3416	−7743	−12441	−810	−397111
SASA/VOL (10 ns)	0.34	0.09	1.34	91	32.3	27.4	70.3	−4240	−7059	−934	−747	−7534
SASA (10 ns)	0.32	0.09	1.36	90	33.3	26.8	70.9	−4232	−7040	−503		−7529
vacuum (100 ns)	0.31	0.07	1.33	93	31.4	22.8	63.3	−4401	−7254			−7870

parametrization[e]							SASA[c]			energies[d]				
$\sigma_{O,N}^{SASAf}$	σ_{C}^{SASAf}	σ^{VOLf}	rmsd[a]	rmsf[a]	R_{gyr}[a]	% H-bonds[b]	phob	phil	total	vdw(u:u)	ele(u:u)	sasa	vol	totpot
−30	5	−80	0.28	0.07	1.35	91	31.9	26.2	68.4	−4258	−6937	−627	−604	−7394
−30	5	−100	0.29	0.07	1.35	91	31.6	25.2	67.2	−4267	−7032	−599	−763	−7524
−30	5	−120	0.29	0.06	1.34	90	31.2	25.5	66.6	−4243	−7134	−609	−929	−7593
−40	5	−80	0.26	0.07	1.38	90	33.4	27.8	72.1	−4156	−7009	−943	−582	−7396
−40	5	−100	0.29	0.07	1.35	90	32.3	27.7	70.9	−4184	−6983	−945	−725	−7445
−40	5	−120	0.24	0.08	1.36	91	32.3	27.7	70.6	−4164	−6945	−948	−904	−7306
−50	5	−80	0.26	0.07	1.40	88	34.4	29.8	76.1	−4113	−6840	−1317	−553	−7155
−50	5	−100	0.25	0.07	1.37	88	32.5	28.5	72.8	−4202	−6879	−1262	−722	−7267
−50	5	−120	0.23	0.07	1.37	89	32.3	29.3	73.2	−4153	−6866	−1299	−874	−7231

[a−f] See footnote to Table 3 for explanations of the quantities described here.

The lengths of the explicit water simulations were trp (1L2Y) and drk (2A36), 20 ns; ubq (1UBQ), if3c (1TIG), lys (1AKI), and talin, 10 ns. The initial parametrization runs using the SASA/VOL model lasted 500 ps; later runs with refined values of σ_i^{SASA} and σ_i^{VOL} lasted 10 ns, as did the SASA simulations. Structures were saved for analysis every 2500 MD steps (5 ps).

RESULTS AND DISCUSSION

Selection of $A_i^{s_1}$ and $A_i^{s_2}$ Values. To keep the model as simple as possible, the same switching function was used for all atom types, thus only a single value of each of $A_i^{s_1}$ and $A_i^{s_2}$ was required. A^{s_1} and A^{s_2} were set to span 1/100 of the maximum atomic SASA found in the six test proteins (1.45 nm^2) so that A^{s_1} = 0.01 nm^2 and A^{s_2} = 0.02 nm^2. Any atom with $A_i < A^{s_2}$ is considered to be

dx.doi.org/10.1021/jp2017117 |J. Phys. Chem. B 2011, 115, 4547–4557

Table 8. Calculated Properties for the talin (189 residues) NMR Model Structure and for Simulations in Explicit Water, in Vacuum, and with the SASA and SASA/VOL Implicit Solvation Models

					SASA[c]			energies[d]				
references	rmsd[a]	rmsf[a]	R_{gyr}[a]	% H-bonds[b]	phob	phil	total	vdw(u:u)	ele(u:u)	ele(u:v)	vdw(u:v)	totpot
NMR			1.93	100	59.8	44.4	116.5	−4966	−5720			−9124
H_2O (10 ns)	0.48	0.19	1.92	96	67.4	47.8	129.7	−4441	−8838	−24007	−976	−919005
SASA/VOL (10 ns)	0.47	0.13	1.87	96	61.4	43.7	116.8	−5625	−6920	−1443	−835	−7787
SASA (10 ns)	0.47	0.16	1.86	95	62.6	41.9	116.0	−5684	−6849	−734		−7761
vacuum (40 ns)	0.43	0.14	1.85	100	60.5	33.8	104.1	−5809	−7129			−8146

parametrization[e]							SASA[c]			energies[d]				
$\sigma_{O,N}^{SASA}$[f]	σ_{C}^{SASA}[f]	σ^{VOL}[f]	rmsd[a]	rmsf[a]	R_{gyr}[a]	% H-bonds[b]	phob	phil	total	vdw(u:u)	ele(u:u)	sasa	vol	.
−30	5	−80	0.36	0.07	1.88	96	60.6	43.0	115.3	−5613	−6818	−986	658	7706
−30	5	−100	0.39	0.07	1.90	96	60.9	42.1	114.7	−5622	−6873	−957	−841	7721
−30	5	−120	0.29	0.06	1.90	97	59.9	41.3	112.1	−5677	−6833	−939	−1038	7772
−40	5	−80	0.31	0.06	1.90	96	62.1	44.7	118.9	−5612	−6671	−1477	−655	7540
−40	5	−100	0.34	0.07	1.89	96	62.7	43.4	118.5	−5628	−6703	−1428	−824	7628
−40	5	−120	0.36	0.07	1.89	96	61.1	43.8	116.6	−5624	−6735	−1445	−1011	7609
−50	5	−80	0.34	0.09	1.90	94	63.5	46.2	112.5	−5539	−6653	−1992	−618	7407
−50	5	−100	0.39	0.08	1.90	94	63.3	46.6	122.9	−5560	−6716	−2013	−803	7483
−50	5	−120	0.38	0.06	1.89	96	60.9	44.7	117.6	−5626	−6744	−1932	−1008	−7572

[a−f] See footnote to Table 3 for explanations of the quantities described here.

buried in the interior of the protein and thus contributes to the volume term; correspondingly, any atom with $A_i > A^{s_2}$ is considered to be solvent exposed and contributes to the SASA term instead.

Tuning of σ_i^{SASA} and σ_i^{VOL}. The form of the SASA component of the solvation term is unchanged from that previously published.[16] The atom-specific σ_i^{SASA} values are limited to just two nonzero values, for hydrophilic ($\sigma_{O,N}^{SASA}$) and hydrophobic (σ_{C}^{SASA}) residues. The contribution to the SASA energy of hydrogen and sulfur atoms is ignored ($\sigma_{H,S}^{SASA} = 0$ kJ·mol^{-1}·nm^{-2}). The σ_i^{SASA} values were extensively tested during the parametrization of the SASA implicit solvation model, resulting in values of $\sigma_{O,N}^{SASA} = -25$ kJ·mol^{-1}·nm^{-2} and $\sigma_{C}^{SASA} = 5$ kJ·mol^{-1}·nm^{-2}.

A negative σ_i^{VOL} is appropriate, as the van der Waals interactions that it represents are favorable. This will encourage the burial of atoms, thus tending to decrease the total SASA. To offset this, the σ_i^{SASA} values must be adjusted to further favor the exposure of atoms to the solvent. This can be achieved by making σ_{C}^{SASA} less positive or, as was done here, making $\sigma_{O,N}^{SASA}$ more negative.

An initial set of simulations scanning σ_i^{VOL} values of several orders of magnitude (−1000−0 kJ·mol^{-1}·nm^{-3}) with the original σ_i^{SASA} values showed that $\sigma_i^{VOL} \sim -100$ kJ·mol^{-1}·nm^{-3} gives energies of a suitable magnitude (data not shown). To maintain the simplicity of the model, just one σ_i^{VOL} value was used for all atom types, termed hereafter simply σ^{VOL}.

Further tests were then carried out in which both the hydrophilic $\sigma_{O,N}^{SASA}$ and σ^{VOL} were tuned simultaneously. The results of these tests are shown in Tables 3−8 and Figure 1.

Because a SASA-based model, by its nature, is directly influenced by and directly affects the SASA, this was the primary criterion by which the SASA/VOL parameters were chosen. In Tables 3−8 and Figure 1, the SASA of the hydrophobic and hydrophilic atoms is shown for each of the tested parameter combinations. All components of the SASA decrease as σ^{VOL} becomes more negative and increase as $\sigma_{O,N}^{SASA}$ becomes more negative. This is just as expected: a more negative σ^{VOL} favors structures with more buried residues, whereas a more negative $\sigma_{O,N}^{SASA}$ favors structures with more solvent-exposed hydrophilic residues. The exposure of more hydrophilic residues also exposes more hydrophobic residues as these are interspersed among the hydrophilic residues and the penalty for exposing hydrophobic residues is small compared to the gain from exposing hydrophilic residues. The effect of different combinations of $\sigma_{O,N}^{SASA}$ and σ^{VOL} values on the SASA also depends on the protein. In general, $\sigma_{O,N}^{SASA} = -40$ or -50 kJ·mol^{-1}·nm^{-2} and $\sigma^{VOL} = -80$ or -100 kJ·mol^{-1}·nm^{-3} produce SASA values between those of the X-ray/NMR model structures and the average over the explicit water simulations, which is the result aimed for.

The R_{gyr} responds to the changes in parameters similarly to the SASA. The decrease in the R_{gyr} when the SASA decreases occurs because the structures with more buried residues are more compact. More expanded structures have fewer hydrogen bonds. In general, the average percentage of hydrogen bonds is lower than in the X-ray structures because these latter structures tend to have a larger number of hydrogen bonds than structures in solution. For all parameter combinations, the hydrogen bonds are maintained similarly to in simulations in explicit water, indicating that the overall organization of the structure remains intact.

As a further measure of the overall stability of the protein structure in each solvent model, the C_α atom-positional rmsd values with respect to the energy-minimized X-ray or NMR model structure and the C_α atom-positional rmsf values were also computed (Tables 3−8). For natively folded and stable proteins at 298 K, such as those studied here, the rmsd and rsmf, particularly for residues in areas of secondary structure, are not expected to become too large. The values obtained for the explicit water simulations were taken as the ideal values for these observables. For the proteins with the largest SASA to volume ratio, namely the smaller proteins and the large but elongated

dx.doi.org/10.1021/jp2017117 |J. Phys. Chem. B 2011, 115, 4547–4557

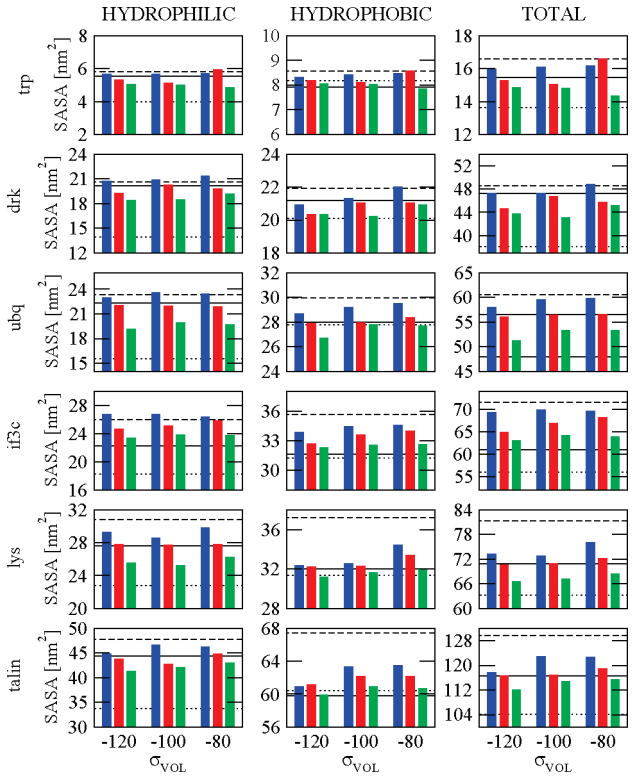


Figure 1. Total and hydrophilic and hydrophobic contributions to the solvent accessible surface area (SASA) for different combinations of σ^{VOL} (in kJ·mol^{-1}·nm^{-3}) as labeled and $\sigma_{O,N}^{SASA}$ = −50 (blue), −40 (red), and −30 (black) (in kJ·mol^{-1}·nm^{-2}) for each of the six test proteins. σ_{C}^{SASA} = 5 kJ·mol^{-1}·nm^{-2} as in the original SASA implicit solvent model.[16] For the simulations using the SASA/VOL model, each bar shows the average over the last 300 ps of a 500 ps simulation. The SASA values of the X-ray or NMR model structure (solid line) or averaged over the entire vacuum (dotted line) and explicit water simulations (dashed line) are also shown.

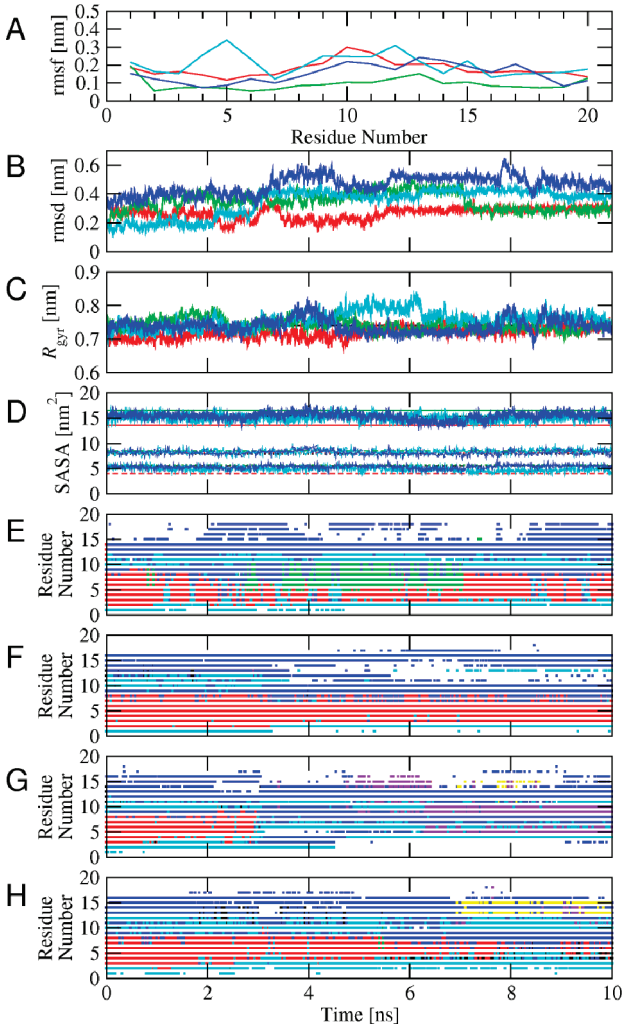


Figure 2. Structural properties of trp. (A) C_α atom-positional root-mean-square fluctuation (rmsf), (B) C_α atom-positional root-mean-square deviation (rmsd) from the minimized NMR model structure, (C) radius of gyration (R_{gyr}), (D) total (solid), hydrophilic (dashed), and hydrophobic (dotted) solvent accessible surface areas (SASA) and (E−H) secondary structure according to the DSSP program:[43] 3_{10}-helix (black), α-helix (red), π-helix (green), bend (blue), β-bridge (yellow), β-strand (violet), and turn (cyan) for simulations (E) in explicit water, (F) using the SASA/VOL solvation term, (G) using the SASA solvation term, and (H) in vacuum. In (A−D), the colors correspond to the NMR model structure (black), simulations in explicit water (green) and in vacuum (red), using the original SASA solvation model with $\sigma_{O,N}^{SASA}$ = −25 kJ·mol^{-1}·nm^{-2} and σ_{C}^{SASA} = 5 kJ·mol^{-1}·nm^{-2} (cyan) and using the SASA/VOL solvation term with $\sigma_{O,N}^{SASA}$ = −40 kJ·mol^{-1}·nm^{-2}, σ_{C}^{SASA} = 5 kJ·mol^{-1}·nm^{-2} and σ_{VOL} = −100 kJ·mol^{-1}·nm^{-3} (blue). Only non-hydrogen atoms contribute to the SASA. In (B) and (C), time series are shown for all simulations, whereas in (D) the average over the simulation is shown other than for the SASA/VOL simulation for which the time series is plotted. The first 10 ns of each simulation is shown for ease of comparison.

talin, the rmsd values obtained using the SASA/VOL solvation term are routinely smaller than those observed in the explicit water simulations, and much less than those of the vacuum and SASA simulations. The rmsd values of if3c and lys are more similar to the values produced by the explicit water simulations, but still smaller than those of the vacuum and SASA simulations.

The rmsf values are very similar for all combinations of σ_i^{SASA} and σ^{VOL} for all proteins. Due to the much shorter length (0.5 ns) of the parametrization simulations, they are significantly smaller than those of the explicit water, SASA, and vacuum simulations but increase in the longer (10 ns) runs using the optimized SASA/VOL parameters to close to the values for the explicit water simulations, especially for the smaller proteins and for talin (Tables 3−8 and Figures 2−7A). Overall, the rmsd and rmsf show that, at least within the first ns of simulation, the addition of the SASA/VOL implicit solvation term maintains the starting structure to a similar degree as explicit solvent molecules, better than the original SASA model and, importantly, better than in a vacuum simulation.

Ideally, in addition to such broad structural features, the solvation free enthalpies ΔG_{solv}, enthalpies ΔH_{solv}, and entropies ΔS_{solv} obtained from implicit and explicit solvation simulations would be compared. This is not straightforward, however, due to the extremely long time scales required to converge the solvent (see Supporting Information for further details). Thus the differences between the energies calculated from implicit and explicit solvent simulations are mostly not discussed here, although they are given in Tables 3−8 for completeness. Note that although the SASA and VOL contributions to the potential energy are listed below the electrostatic and van der Waals contributions to the potential energy

dx.doi.org/10.1021/jp2017117 |J. Phys. Chem. B 2011, 115, 4547–4557

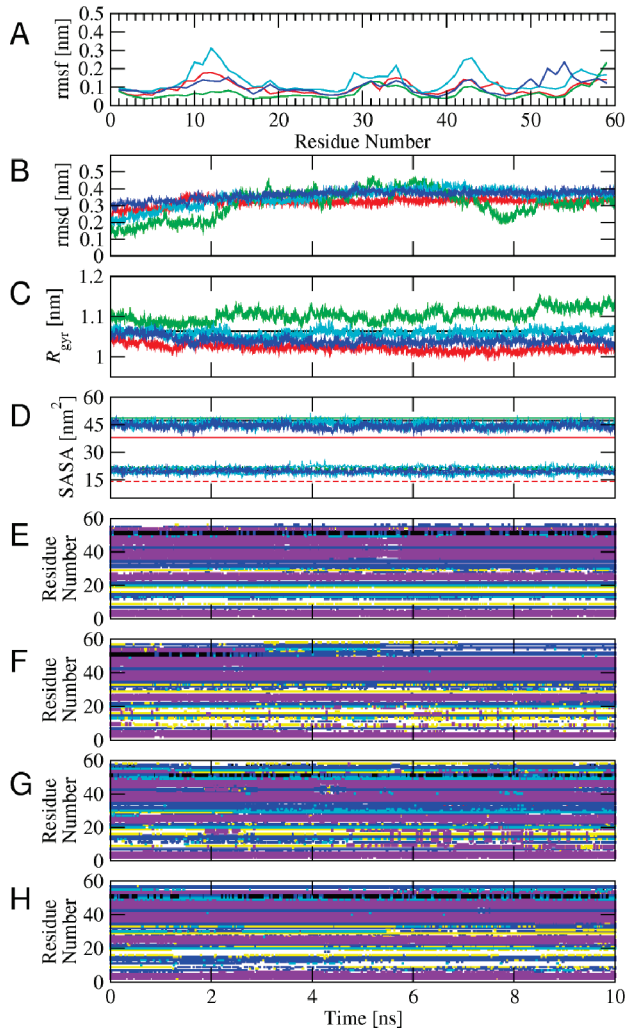


Figure 3. Structural properties of drk. See caption to Figure 2 for explanations of the quantities that are displayed.

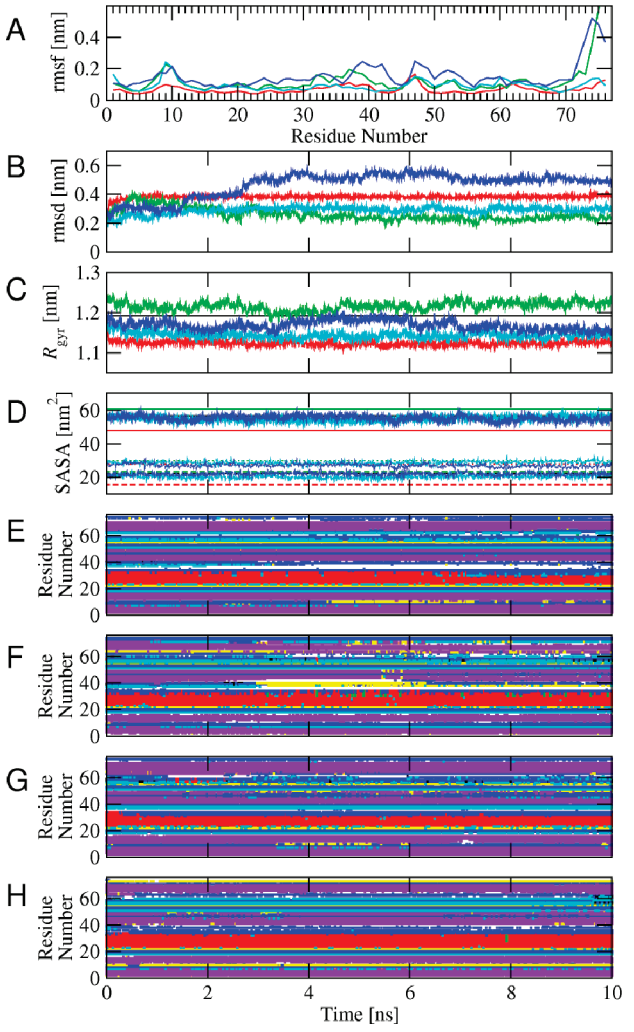


Figure 4. Structural properties of ubq. See caption to Figure 2 for explanations of the quantities that are displayed. The reference structure was the X-ray structure.

of the explicit water simulations, they do not correspond directly. Both the SASA and VOL terms encompass van der Waals and electrostatic effects, the latter of which are reduced due to the neutralization of the charged moieties in the solute. The SASA and VOL energies respond in the expected manner to the changes in $\sigma_{O,N}^{SASA}$ and σ^{VOL}.

The complex interdependence of the effects of the values of σ_i^{SASA} and σ^{VOL} with the size and shape of the protein complicated somewhat the choice of the optimal values of σ_i^{SASA} and σ^{VOL}. After scrutiny of the variety of data presented in Tables 3−8 and Figure 1, values of $\sigma_{O,N}^{SASA} = -40$ kJ·mol^{-1}·nm^{-2} and $\sigma^{VOL} = -100$ kJ·mol^{-1}·nm^{-3} were chosen, as for the proteins considered here, these provided the best compromise in terms of matching the R_{gyr}, hydrogen bonds, and in particular, the SASA to the values of these quantities as obtained from the X-ray or NMR model structures and the averages from the explicit water simulations (which do not necessarily agree with each other) alongside a reasonable C_α atom-positional rmsd with respect to the energy-minimized X-ray or NMR model structures.

Optimized Parameters. To provide a more stringent test, longer, 10 ns simulations of each of the six test proteins using the optimized values of σ_C^{SASA}, $\sigma_{O,N}^{SASA}$, and σ^{VOL} were run. The results of these are summarized in Tables 3−8 and Figures 2−7. The simulation time of 10 ns was judged to be sufficient based on the convergence of the C_α atom-positional rmsf values (Supporting Information).

The total, hydrophilic and hydrophobic SASA in the SASA/VOL and SASA simulations are close to the values of the X-ray or NMR model structures. In explicit water, all components of the SASA are larger than in the X-ray or NMR model structure, whereas in vacuum, the hydrophilic SASA, in particular, is smaller than that of the reference structure, because the favorable interactions of the polar residues with the solvent are missing. This is alleviated by the SASA/VOL model, which, despite the compromise inherent in the choices of the solvation parameters $\sigma_{O,N}^{SASA}$ and σ^{VOL}, is able to maintain the SASA of the initial structures in almost all cases.

The secondary structure of all proteins is mostly stable and similar for all four solvent models. In fact, the SASA/VOL simulation is the only one in which the secondary structure of trp is maintained. For this small protein, all buried residues are close to the solvent, making the volume contribution particularly important.

The C_α atom-positional rmsf is uniformly small (<0.2 nm) throughout the residue sequence for all solvent models, other than for residues located in loops and turns (Figures 2−7A) and for trp, for which the larger rmsf reflects the less-structured

dx.doi.org/10.1021/jp2017117 |J. Phys. Chem. B 2011, 115, 4547–4557

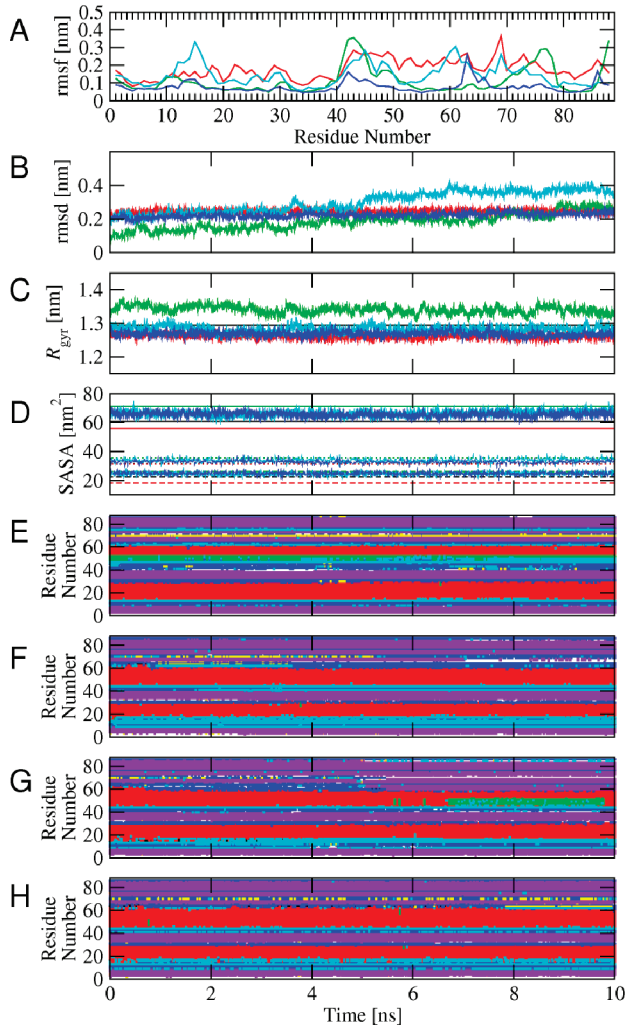


Figure 5. Structural properties of if3c. See caption to Figure 2 for explanations of the quantities that are displayed. The reference structure was the X-ray structure.

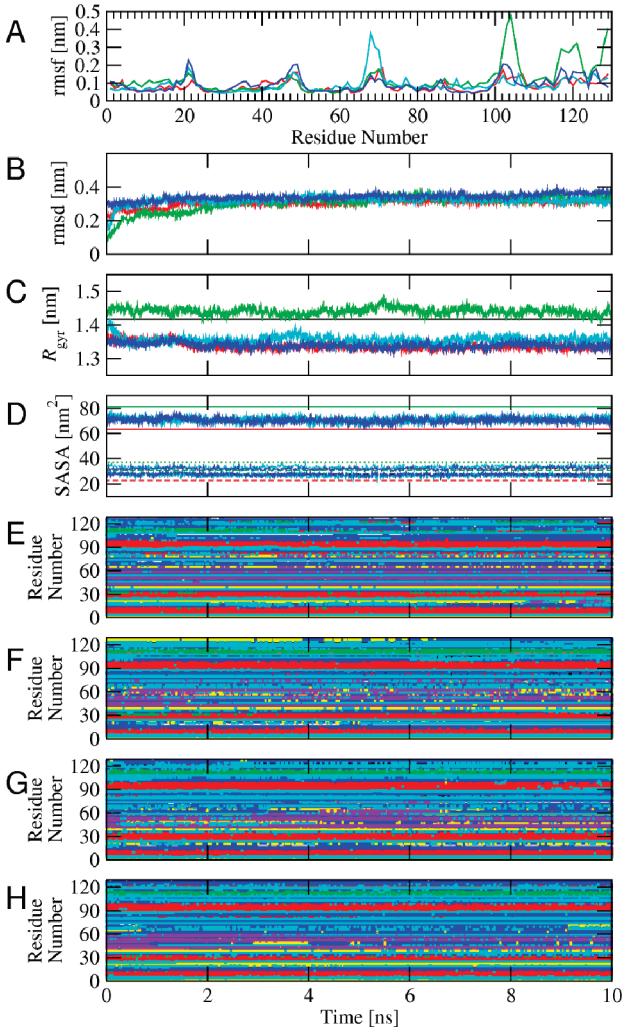


Figure 6. Structural properties of lys. See caption to Figure 2 for explanations of the quantities that are displayed. The reference structure was the X-ray structure.

nature of this small protein. In general, the rmsf is smallest for the vacuum simulations. With the SASA/VOL solvent model, the residues with the largest rmsf are mostly the same as in the explicit water simulations, whereas with the original SASA model, this is not necessarily the case. For the larger proteins (if3c, lys, and talin), the height of the peaks in the rmsf is greatest for the explicit water simulations, whereas for drk and ubq, they are higher with the SASA and SASA/VOL models, respectively. This is likely due to an overestimation of the volume correction for the proteins, for which a greater proportion of the buried residues are too far from the solvent to have significant interactions.

The C_α atom-positional rmsd of each protein is in general larger in the SASA, SASA/VOL, and explicit water simulations than in vacuum. Other than a few exceptions, the rmsd with the SASA/VOL model is similar to or less than that of the simulations using explicit water, indicating that the protein structure is stable in the revised implicit solvent model.

In general, the R_{gyr} in vacuum is significantly lower than that of the X-ray or NMR model structure, and the R_{gyr} in explicit water is larger, with those of the SASA and SASA/VOL models lying somewhere in between. For trp and ubq, the R_{gyr} in the SASA/VOL solvent model is similar to that of the NMR and X-ray structures, respectively, but for drk, if3c, and lys it is more like that of the vacuum simulation. The larger R_{gyr} of the explicit water simulations is related to the mostly larger rmsf of the C-termini and loops; the fluctuation of these regions tends to move them away from the remainder of the protein, increasing the R_{gyr}, whereas in vacuum and, for the most part, in simulations with the SASA/VOL model, they remain close to the protein core.

To illustrate the large savings in computational cost relative to simulations using explicit water, the time taken and speedup for 20 ps MD simulations in vacuum and using the SASA or SASA/VOL implicit solvent models was computed for each of the six model proteins considered here (Table 9). The speedup is in general more significant for smaller proteins, as the box size is determined by enforcing a minimum distance between solute atoms and the edge of the box, meaning that for smaller proteins, solvent molecules constitute a larger proportion of the system. The speedup also depends on the shape of the protein and thus of the box. The SASA and SASA/VOL models offer mostly very similar speedups, thus the increased physical accuracy obtained by adding the volume term comes at essentially no additional cost. While the simulations in vacuum of course proceed even

dx.doi.org/10.1021/jp2017117 |J. Phys. Chem. B 2011, 115, 4547–4557

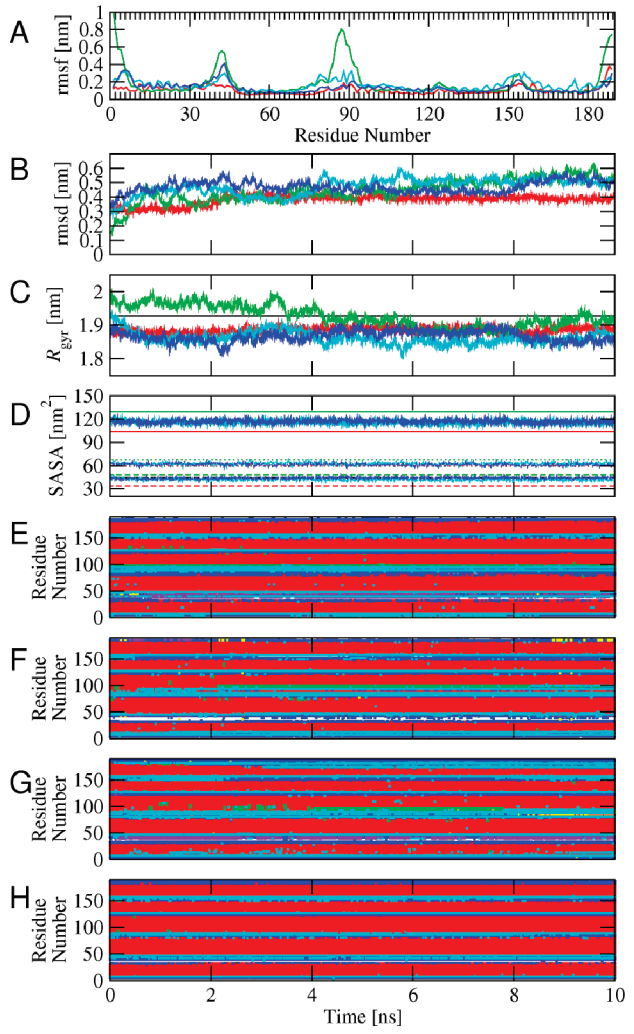


Figure 7. Structural properties of talin. See caption to Figure 2 for explanations of the quantities that are displayed.

faster, the complete neglect of solvation effects means that the structures sampled may not be physically relevant.

Although the chosen σ_C^{SASA}, $\sigma_{O,N}^{SASA}$, and σ^{VOL} parameter values were a compromise, rather than optimized for each protein, overall, the SASA, rmsf, and secondary structure are similar to those of the explicit water simulations, and the differences in R_{gyr} can be accounted for by loop and C-terminal flexibility. The unexpectedly large rmsf of certain residues observed in the SASA solvent model are eliminated in the SASA/VOL model. Additionally, the R_{gyr} is in general slightly larger in the SASA/VOL model than in the original SASA model, making it more like that observed in the explicit water simulations than in the vacuum simulations. Thus the addition of the volume term provides an improvement on the original SASA model, particularly for small proteins for which the majority of the buried residues are close enough to the surface to have significant interactions with the solvent. We note, however, that for proteins of spherical shape that are considerably larger than the ones studied here, the interior volume that is far from the surface will become non-negligible. Since in the SASA/VOL model, all interior atoms contribute to the solvation energy term $V_{solv}^{VOL}(\mathbf{r}^N)$, but in reality, the atoms that are far from the surface will have a negligible van der Waals interaction with the solvent, the σ^{SASA} and σ^{VOL} parameter values will have to be tested and perhaps recalibrated if very large, spherical proteins are to be simulated using the SASA/VOL model. Overall, the similar behavior of most of the key structural properties in SASA/VOL and in explicit water simulations over nanosecond time-scales confirms that the SASA/VOL model provides a reasonable substitute for explicit water, allowing simulations to be conducted at greater speed without too much distortion.

Table 9. Time Taken and Speedup Relative to Explicit Water for MD Simulations[a] Using the SASA/VOL and SASA Implicit Solvation Models and in Vacuum

	H_2O[b]	SASA/VOL		SASA		vacuum	
protein	time[c]	time	speedup	time	speedup	time	speedup
trp	1467	17	86	16	90	6	240
drk	3408	125	27	126	27	23	142
ubq	5652	197	29	195	29	30	183
if3c	6110	270	23	268	23	37	165
lys	11298	568	20	563	20	62	182
talin	21854	972	22	905	24	74	295

[a] 20 ps benchmark simulations were run using GROMOS MD++ compiled using gcc 4.4.5 on a workstation class PC running Ubuntu Linux 10.10. Hardware specification: Intel Core i7 CPU 920 (2.67 GHz clock, 8092 KB cache), 6 GB RAM, ASUS P6T Deluxe V2 motherboard. Only one of the four cores was used (no parallelization). [b] H_2O refers to explicit (SPC) water. [c] The time is in seconds.

CONCLUSIONS

The previously published SASA implicit solvent model has been extended to include the effects of buried volume. Reoptimization of the model parameters resulted in a set of parameters for which the key structural properties of six test proteins are maintained better than or similarly to with the original SASA model throughout a 10 ns simulation. Because the interior volume contribution is calculated using the atomic surface areas, there is little additional cost. Thus the SASA/VOL solvation model offers a fast alternative to doing explicit water simulations in cases where it is important to maintain the average effects of solvation. In its current parametrization, it is most appropriate for not too large spherical or large nonspherical proteins, where a significant proportion of the buried residues lie close to the surface.

ASSOCIATED CONTENT

Supporting Information. Details of the simulation methods, spatial derivatives with respect to atomic coordinates of the mean solvation potential energy term, decomposition of the solvation free enthalpy, and figure showing the behavior of the C_α atom-positional rmsf values over time. This material is available free of charge via the Internet at http://pubs.acs.org/.

AUTHOR INFORMATION

Corresponding Author

*E-mail: wfvgn@igc.phys.chem.ethz.ch. Phone: +41 44 632 5501. Fax: +41 44 632 1039.

ACKNOWLEDGMENT

We are grateful to Jens Kleinjung, Bruno Horta, and Philippe Hünenberger for many stimulating discussions. This work was

supported by the National Center of Competence in Research (NCCR) in Structural Biology, by Grant No. 200020-121913 of the Swiss National Science Foundation, by grant number 228076 of the European Research Council and by the ETH Research Grant ETH-01 09-2.

■ ABBREVIATIONS

CASA, Coulomb/accessible surface area; MD, molecular dynamics; NMR, nuclear magnetic resonance; PDB, protein data bank; R_{gyr}, radius of gyration; rmsd, root-mean-square deviation; rmsf, root-mean-square fluctuation; SASA, solvent accessible surface area; SPC, simple point charge; VOL, (buried) volume

■ REFERENCES

(1) Daura, X.; Mark, A. E.; van Gunsteren, W. F. *Comput. Phys. Commun.* **1999**, *123*, 97–102.
(2) van Gunsteren, W. F.; Luque, F. J.; Timms, D.; Torda, A. E. *Annu. Rev. Biophys. Biomol. Struct.* **1994**, *23*, 847–63.
(3) Roux, B.; Simonson, T. *Biophys. Chem.* **1999**, *78*, 1–20.
(4) Feig, M.; Brooks, C. L. *Curr. Opin. Struct. Biol.* **2004**, *14*, 217–24.
(5) Chen, J.; Brooks, C. L.; Khandogin, J. *Curr. Opin. Struct. Biol.* **2008**, *18*, 140–8.
(6) Chothia, C. *Nature* **1974**, *248*, 338–9.
(7) Eisenberg, D.; McLachlan, A. D. *Nature* **1986**, *319*, 199–203.
(8) Ooi, T.; Oobatake, M.; Némethy, G.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S. A.* **1987**, *84*, 3086–90.
(9) Kang, Y. K.; Gibson, K. D.; Némethy, G.; Scheraga, H. A. *J. Phys. Chem.* **1988**, *92*, 4739–42.
(10) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. *J. Am. Chem. Soc.* **1990**, *112*, 6127–9.
(11) Sharp, K.; Nicholls, A.; Fine, R.; Honig, B. *Science* **1991**, *252*, 106–9.
(12) Vila, J.; Williams, R. L.; Vásquez, M.; Scheraga, H. A. *Proteins: Struct., Funct., Genet.* **1991**, *10*, 199–218.
(13) Schiffer, C. A.; Caldwell, J. W.; Stroud, R. M.; Kollman, P. A. *Protein Sci.* **1992**, *1*, 396–400.
(14) Wesson, L.; Eisenberg, D. *Protein Sci.* **1992**, *1*, 227–35.
(15) Stouten, P.; Frömmel, C.; Nakamurah, H.; Sander, C. *Mol. Simul.* **1993**, *10*, 97–120.
(16) Fraternali, F.; van Gunsteren, W. F. *J. Mol. Biol.* **1996**, *256*, 939–48.
(17) Lazaridis, T.; Karplus, M. *Proteins: Struct. Funct. Genet.* **1999**, *35*, 133–52.
(18) Ferrara, P.; Apostolakis, J.; Caflisch, A. *Proteins: Struct., Funct., Genet.* **2002**, *46*, 24–33.
(19) Fraternali, F.; Cavallo, L. *Nucleic Acids Res.* **2002**, *30*, 2950–60.
(20) Cerutti, D. S.; Ten Eyck, L. F.; McCammon, J. A. *J. Chem. Theory Comput.* **2004**, *1*, 143–52.
(21) Lopes, A.; Alexandrov, A.; Bathelt, C.; Archontis, G.; Simonson, T. *Proteins: Struct., Funct., Bioinf.* **2007**, *67*, 853–67.
(22) Kang, Y. K.; Nemethy, G.; Scheraga, H. A. *J. Phys. Chem.* **1987**, *91*, 4105–9.
(23) Lum, K.; Chandler, D.; Weeks, J. D. *J. Phys. Chem. B* **1999**, *103*, 4570–7.
(24) Hummer, G.; Garde, S.; García, A. E.; Pratt, L. R. *Chem. Phys.* **2000**, *258*, 349–70.
(25) Huang, D. M.; Chandler, D. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 8324–7.
(26) Rajamani, S.; Truskett, T. M.; Garde, S. *Proc. Natl. Acad. Sci. U. S.A.* **2005**, *102*, 9475–80.
(27) Dzubiella, J.; Swanson, J. M. J.; McCammon, J. A. *J. Chem. Phys.* **2006**, *124*, 084905–12.
(28) Floris, F.; Tomasi, J. *J. Comput. Chem.* **1989**, *10*, 616–27.
(29) Floris, F. M.; Tomasi, J.; Ahuir, J. L. P. *J. Comput. Chem.* **1991**, *12*, 784–91.
(30) Gallicchio, E.; Kubo, M. M.; Levy, R. M. *J. Phys. Chem. B* **2000**, *104*, 6271–85.
(31) Pitera, J. W.; van Gunsteren, W. F. *J. Am. Chem. Soc.* **2001**, *123*, 3163–4.
(32) Zacharias, M. *J. Phys. Chem. A* **2003**, *107*, 3000–4.
(33) Gallicchio, E.; Levy, R. M. *J. Comput. Chem.* **2004**, *25*, 479–99.
(34) Choudhury, N.; Pettitt, B. M. *J. Am. Chem. Soc.* **2005**, *127*, 3556–7.
(35) Malham, R.; Johnstone, S.; Bingham, R. J.; Barratt, E.; Phillips, S. E. V.; Laughton, C. A.; Homans, S. W. *J. Am. Chem. Soc.* **2005**, *127*, 17061–7.
(36) Wagoner, J. A.; Baker, N. A. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 8331–6.
(37) Christen, M.; Hünenberger, P. H.; Bakowies, D.; Baron, R.; Bürgi, R.; Geerke, D. P.; Heinz, T. N.; Kastenholz, M. A.; Kräutler, V.; Oostenbrink, C.; Peter, C.; Trzesniak, D.; van Gunsteren, W. F. *J. Comput. Chem.* **2005**, *26*, 1719–51.
(38) van Gunsteren,W. F.; Billeter, S.; Eising, A.; Hünenberger, P.; Krüger, P.; Mark, A.; Scott,W.; Tironi, I. *Biomolecular simulation: the GROMOS96 manual and user guide*; Vdf Hochschul-verlag AG an der ETH Zürich: Zürich, Switzerland, 1996.
(39) Hasel, W.; Hendrickson, T. F.; Still, W. C. *Tetrahedron Comput. Methodol.* **1988**, *1*, 103–16.
(40) Wodak, S. J.; Janin, J. *Proc. Natl. Acad. Sci. U.S.A.* **1980**, *77*, 1736–40.
(41) Schuler, L. D.; Daura, X.; van Gunsteren, W. F. *J. Comput. Chem.* **2001**, *22*, 1205–18.
(42) Oostenbrink, C.; Villa, A.; Mark, A. E.; van Gunsteren, W. F. *J. Comput. Chem.* **2004**, *25*, 1656–76.
(43) Kabsch, W.; Sanders, C. *Biopolymers* **1983**, *22*, 2577–637.