Converging Free Energy Estimates: MM-PB(GB)SA Studies on the Protein–Protein Complex Ras–Raf

HOLGER GOHLKE, DAVID A. CASE
Department of Molecular Biology, The Scripps Research Institute, 10550 N. Torrey Pines Rd., La Jolla, California 92037

Received 8 May 2003; Accepted 2 September 2003

Abstract: Estimating protein–protein interaction energies is a very challenging task for current simulation protocols. Here, absolute binding free energies are reported for the complex H-Ras/C-Raf1 using the MM-PB(GB)SA approach, testing the internal consistency and model dependence of the results. Averaging gas-phase energies (MM), solvation free energies as determined by Generalized Born models (GB/SA), and entropic contributions calculated by normal mode analysis for snapshots obtained from 10 ns explicit-solvent molecular dynamics in general results in an overestimation of the binding affinity when a solvent-accessible surface area-dependent model is used to estimate the nonpolar solvation contribution. Applying the sum of a cavity solvation free energy and explicitly modeled solute–solvent van der Waals interaction energies instead provides less negative estimates for the nonpolar solvation contribution. When the polar contribution to the solvation free energy is determined by solving the Poisson–Boltzmann equation (PB) instead, the calculated binding affinity strongly depends on the atomic radii set chosen. For three GB models investigated, different absolute deviations from PB energies were found for the unbound proteins and the complex. As an alternative to normal-mode calculations, quasiharmonic analyses have been performed to estimate entropic contributions due to changes of solute flexibility upon binding. However, such entropy estimates do not converge after 10 ns of simulation time, indicating that sampling issues may limit the applicability of this approach. Finally, binding free energies estimated from snapshots of the unbound proteins extracted from the complex trajectory result in an underestimate of binding affinity. This points to the need to exercise caution in applying the computationally cheaper "one-trajectory-alternative" to systems where there may be significant changes in flexibility and structure due to binding. The best estimate for the binding free energy of Ras–Raf obtained in this study of -8.3 kcal mol^{-1} is in good agreement with the experimental result of -9.6 kcal mol^{-1}, however, further probing the transferability of the applied protocol that led to this result is necessary.

 J Comput Chem 25: 238–250, 2004

Key words: absolute binding free energy; protein–protein interactions; generalized Born model; entropy; Ras

Introduction

Protein–protein interactions form the basis for a wide range of biological and biophysical processes, such as hormone–receptor interactions, proteinase inhibition, antibody–antigene interactions, signal transduction, and enzyme allostery.[1,2] Genome sequencing projects and knowledge of the entire protein repertoires of many organisms provide large-scale information about protein–protein interaction networks.[3,4] Despite an increasing number of crystallographically determined protein–protein complexes, thermodynamic experiments, and mutational studies,[1,5–7] our understanding of the energetics and dynamics of the mutual binding of two proteins on an atomic level is still limited.

In recent years, computational studies have provided an interesting complement to analyses of experimental protein–protein data.[8] Gaining insight into the energetics of binding is a problem that is extremely difficult to attack by conventional computational free energy techniques. Alternative methods range from the application of empirical free energy functions[9–11] and structure-based pair potentials[12,13] to all-atom molecular mechanics methods including explicit[14] or implicit[15–19] representations of the solvent end entropy estimates to predict absolute binding free energies or changes in the binding free energy upon (structural) changes to the

Correspondence to: D. A. Case; e-mail: case@scripps.edu

Contract/grant sponsor: NIH; contract/grant number: RR12255.

This article includes Supplementary Material available from the authors upon request or via the Internet at ftp://ftp.wiley.com/public/journals/jcc/suppmat/25/238 or http://www.interscience.wiley.com/jpages/0192-8651/suppmat/25/v25.238.html

systems. The latter approaches, in particular, allow addressing the origin of binding in terms of contributions from electrostatic and van der Waals interactions and changes in solvation.[8,15,20] Determining (absolute) screened electrostatic, nonpolar, and entropic contributions to binding is still a major challenge in this connection. Here, we investigate different computational alternatives to estimate these contributions in the context of absolute binding free energy calculations for protein–protein complexes.

The system we chose is the complex between human H-Ras and the Ras-binding domain of C-Raf1 (Ras-Raf), which is central to the signal transduction cascade.[21] Binding of GTP or GDP to Ras activates or inactivates the protein, respectively, and induces conformational transitions mainly in two regions that have been called "switch I" (residues 30–37) and "switch II" (residues 60–76). Being activated by "upstream" signals, Ras interacts with "downstream" effectors such as Raf mostly via the "effector region" (residues 32–40).[22] Molecular structures for the unbound proteins[23–26] or complexes closely related to Ras–Raf[27,28] have been determined either by X-ray crystallography or by NMR spectroscopy. They show that the Ras–effector interactions mainly occur due to a formation of an interprotein β-sheet, resulting in an interface size of approx. 1200 Å^2,[21] and only small or moderate changes of the protein structures upon binding. The binding free energy of the effector protein to Ras has been determined by isothermal titration calorimetry to be -9.7 kcal mol^{-1}.[29]

Ras–effector interactions have been repeatedly investigated by means of theoretical methods. Among others, studies included MD simulations on Ras or Rap-1A bound to Raf aimed at identification of effector domains in the Raf protein[30,31] or only of Raf to investigate conformational effects of the R89K mutation.[32] Zeng et al.[33] performed alchemical free energy calculations on the R89K mutant of Raf, while Muegge et al.[17] concentrated mainly on electrostatic contributions to the Rap-1A–Raf interaction. Recently, we studied structural determinants of the binding free energy of Ras–Raf and Ras–Ra1GDS by means of free energy decomposition.[34]

(Absolute) binding free energies for Ras–Raf are computed here as the sum of gas-phase energies, solvation free energies, and entropic contributions, averaged over snapshots from molecular dynamics trajectories. This so-called MM-PB(GB)SA approach (MM: Molecular Mechanics; PB: Poisson–Boltzmann; GB: Generalized Born; SA: Surface Area)[16,35] has been applied to compare relative stabilities of different conformations of nucleic acids,[35] to identify correctly folded proteins,[36] and to estimate binding affinities of small molecules binding to proteins.[37–39] It has also been used to predict the effects of amino acid mutations on binding affinities.[15,18,40] More recently, electrostatic contributions to solvation free energies have also been calculated by generalized Born models in this context.[35,41,42] Because GB is significantly faster than PB, the former is an attractive alternative to the latter. Modifications of GB have been reported for molecules with larger interior regions.[43,44] Often, however, GB models have been parameterized primarily to reproduce energy differences between different conformations of one molecule in agreement with results from Poisson(–Boltzmann) calculations.[41,43] Here, we test the performance of GB models in the case of binding free energy calculations, which requires giving accurate solvation free energies across different molecular species. Recently, deficiencies of solvent-accessible surface area-dependent models to estimate the total nonpolar solvation contribution in the case of macromolecules have been discussed.[45–49] These deficiencies mainly arise due to missing solvation contributions of interior (buried) atoms. An alternative model for nonpolar solvation uses a cavity solvation free energy term plus solute–solvent van der Waals interactions.[49–51] We compare both models here with respect to their outcome on the calculated binding free energy.

In this study, three 12-ns explicit solvent MD simulations were carried out for the proteins Ras, Raf, and the Ras–Raf complex. The absolute binding free energy of Ras–Raf complex formation was estimated from energetic and entropic contributions, calculated for snapshots extracted from the trajectories and averaged, following the MM-PB(GB)SA approach. Internal consistency and model dependence of the calculations were probed in five different ways. First, statistical convergence of average energies was verified. Second, variation of the continuum solvent models resulted in either over- or underbinding of the proteins if PB calculations were used (depending on the atomic radii applied), but overestimation of the binding affinity in the case of GB calculations. Third, taking into account van der Waals interactions between buried solute and solvent atoms, nonpolar desolvation becomes less negative compared to results of a solvent-accessible surface area-dependent model. Fourth, entropy changes of the solute molecules were estimated by normal mode analyses. Alternatively, quasiharmonic analyses were also applied, but converged results were not obtained from simulations as long as 10 ns. Finally, using snapshots of the unbound molecules extracted only from the trajectory of the Ras–Raf complex, the binding affinity was underestimated compared to the case where snapshots were extracted from three different trajectories. Overall, despite considerable methodological progress in recent years, absolute binding free energy calculations for macromolecular systems still present a challenging task.

Methods and Data

Structures from Molecular Dynamics Simulations

The snapshots of the unbound proteins and complexes were taken from molecular dynamics (MD) simulations that have been described recently.[34] In general, all simulations were performed in a consistent fashion with the AMBER 7 suite of programs,[52] together with the Cornell et al.[53] force field and the TIP3P model for water.[54] Bonded parameters for the triphosphate moiety of GTP were taken from Leach and Klein[55] and atomic partial charges for GTP^{4-} were derived using the RESP procedure.[56] Nonbonded parameters for Mg^{2+} were taken from[57] and were adapted to the AMBER combining rules. Periodic boundary conditions were applied, using the particle mesh Ewald (PME) method[58] to treat long-range electrostatic interactions. The sizes of the primary simulation cells were (in Å^3): $55 \times 63 \times 61$ (Raf), $69 \times 67 \times 72$ (Ras), $74 \times 71 \times 88$ (Ras–Raf). Bond lengths involving bonds to hydrogen atoms were constrained using SHAKE.[59] The time step for all MD simulations was 2 fs and a direct-space nonbonded cutoff of 9 Å was used. After minimization, the solvated systems were heated up to 300 K using NVT-MD, and the solvent density was adjusted using NPT-MD. In all cases, 2 ns of unconstrained

NVT-MD at 300 K with a time constant of 2.0 ps for heat bath coupling were performed before snapshots for binding free energy calculations were extracted from production runs of 10-ns length each. For further details concerning the simulation protocols see ref. 34.

The starting structures for the simulations of the human unbound proteins and complexes were taken from the Protein Data Base (PDB codes: 1rrb, 121p, 1gua)[60] and modified to achieve consistency with respect to the biological source and the number of amino acids. Below, structural results of the simulations obtained for snapshots extracted at 5-ps intervals are summarized. More detailed information is published elsewhere.[34]

The time series of the rmsd of backbone atoms from the experimental starting structures show values that vary between 0.7 and 1.8 Å for all systems but Raf (maximum at 3.2 Å). In this case, one has to note, however, that the Raf starting structure was derived from NMR experiments and that the region of loop L1 (residues 101–109) is not well defined in the experimental solution structure. It is also this region, which reveals major conformational changes (4.7 Å) when superimposing average and starting structures of the unbound proteins. In the case of Ras, the largest changes occur—not unexpectedly—in the switch I and switch II regions. Comparing the average structures of bound and unbound proteins, the effector protein shows larger deviations (2.2 Å) than the Ras structure (0.8 Å). Over the course of the MD simulations, the rmsd of the GTP atoms and the one for the Mg^{2+} ion are below 0.8 and 0.4 Å, respectively, with respect to the starting structures, and neither of the two water molecules coordinated to Mg^{2+} exchange with the bulk solvent. This finding, being in agreement with the experiment,[61] justifies considering these two water molecules as part of the solute in the subsequent calculations. One water molecule could be identified that has atomic fluctuations below 1 Å and resides in the interface between Ras and Raf, mediating interactions between both proteins. In the unbound proteins, in contrast, no equally restricted water molecules were found. A survey of hydrogen bonds formed between both proteins reveals two salt bridges in the case of Ras–Raf. Finally, comparing hydrogen bonds found during MD simulations with those from the solution structure of the complex determined by NMR indicates that the interaction pattern across the interface agrees to a large extent (but not perfectly).

For the binding free energy calculations, 500 snapshots of the unbound proteins and the complexes were extracted from 10-ns production runs of separate trajectories at time intervals of 20 ps ("set 1"). Additionally, snapshots of the unbound proteins were also extracted from a single trajectory of the complex ("set 2"). Although the latter alternative is attractive due to its lower computational demand, it is only valid if the intramolecular strain accommodated by the components upon association is negligible or if no significant conformational changes occur. To test the influence of different models for estimating solvation free energies, the first 150 snapshots of "set 1" were used, corresponding to the 3-ns MD trajectories after 2 ns of unrestrained equilibration. This resulted in "set 3." In all cases, all counterions were stripped, as were all water molecules but the two closest to the Mg^{2+} ions in Ras and Ras–Raf.

Calculations of Binding Free Energies

Binding free energies were calculated, following the MM-PB(GB)SA approach,[16,35] according to

$$\Delta G_{\text{binding}} = \langle G^{\text{complex}}(i)\rangle_i - \langle G^{\text{protein1}}(i)\rangle_i - \langle G^{\text{protein2}}(i)\rangle_i. \quad (1)$$

Here, $\langle \cdots \rangle$ denotes an average over snapshots i taken from MD trajectories and $G^x(i)$ was estimated from contributions of gas-phase energies, solvation free energies, and entropies.

$$G^x(i) = H^x_{\text{gas}}(i) + H^x_{trans/\text{rot}} + G^x_{\text{solvation}}(i) - TS^x(i). \quad (2)$$

For snapshots of the unbound proteins taken from a single trajectory of the complex, the free energy of binding was computed as

$$\Delta G_{\text{binding}} = \langle G^{\text{complex}}(i) - G^{\text{protein1}}(i) - G^{\text{protein2}}(i)\rangle_i. \quad (3)$$

"Gas-phase" energies H_{gas} of the solutes were calculated by summing contributions from internal energies (including bond, angle, and torsional angle energies), electrostatic energies, and van der Waals energies using the Cornell et al. force field[53] with no cutoff. In the classical limit, the energy $H^x_{trans/\text{rot}}$ due to six translational and rotational degrees of freedom is $6 * \frac{1}{2}$RT (=1.79 kcal mol^{-1} at 300 K). This contribution has been omitted in earlier MM-PBSA calculations of absolute binding free energies.[39,41] Effects on binding free energy due to small volume changes on conversion from a canonical to an isothermal-isobaric ensemble are neglected.

Solvation Free Energies

Solvation free energies $G^x_{\text{solvation}}(i)$ were computed as sum of a polar and a nonpolar contribution using a continuum representation of the solvent. In this model, a low dielectric solute with embedded charges is surrounded by a high dielectric continuum solvent. If not otherwise stated, a dielectric constant inside and outside the solute of 1 and 80 was applied, respectively. The partial charges on the solute atoms are taken from the Cornell et al.[53] force field to be consistent with the energetics of the explicit solvent simulations used to generate the snapshots.

The polar contribution was calculated either by solving the Poisson equation (PE) in case of zero salt conditions or the linearized and nonlinearized Poisson–Boltzmann (PB) equations for nonzero salt concentrations as implemented in Delphi II.[62] For the calculations, two grid points per Å were used, the solute filled 80% of the grid box, and 500–1000 finite difference iterations were performed to ensure convergence of the results (i.e., the mean absolute change in potential at the grid points was below 10^{-4} kT per unit charge). Potentials at the boundaries of the finite-difference grid were set to the sum of Debye–Hückel potentials and radii of the atoms were either taken from the PARSE parameter set[63] or from Bondi.[64] Alternatively, generalized Born models[65–67] can be considered as an approximation to the computationally intensive PB approach. Recently, modifications of the methods have been proposed that improve their performance with respect to molecules with larger interior regions.[43,44] Hence, we also estimated the

electrostatic component of the solvation free energy by GB models as implemented in AMBER 7.[52] In particular, this includes models by Tsui et al.[41] ("vtGB"), Onufriev et al.[43] ("aoGB"), and Jayaram et al.[68] ("MGB"). In all cases, radii, as reported in the respective studies, were used. To compare electrostatic contributions to the solvation free energy as calculated by both methods, the same radii were used for PE calculations as for the respective GB calculations. In addition, to comply with the way the MGB model was derived,[68] polar contributions as calculated by PE were determined as differences of calculations using dielectric constants for the interior/exterior of the solute of 2/80 and 2/1 in this case; all other calculations used an interior dielectric of 1 and an exterior dielectric of 80.

The dependence of the binding free energy on the Mg^{2+} radius used for continuum solvent calculations has been reported in the context of MM-PBSA.[69] To test the influence on the binding free energy, radii of 0.99 Å,[70] 1.18 Å,[70] and 1.45 Å,[69] were chosen for the Mg^{2+} ion. In our case, applying these three different radii yielded differences in ΔG_{total} of 1 kcal mol^{-1} at maximum for both GB and PB calculations (data not shown). The insensitivity in case of finite-difference calculations to this radius is understandable, because the two water molecules together with the protein atoms completely shield it from the dielectric boundary. The scaling factor[71] for Mg^{2+} for the GB models was set to 0.8. Polar contributions to solvation free energy were computed at 50 mM ionic strength applying the linearized PB model and an extension[72] to the GB model. It has been shown recently[34] that this disfavors binding of the oppositely charged molecules Ras and Raf by approximately 6 kcal mol^{-1} compared to zero salt conditions. PB and GB calculations give very similar trends in that respect (data not shown).

The nonpolar contribution to the solvation free energy due to cavity formation and van der Waals interactions between the solute and the solvent was estimated by a solvent-accessible surface area ($SASA$) dependent term:

$$G^x_{\text{nonpolar}}(i) = \gamma SASA^x(i) + b. \quad (4)$$

The $SASA^x(i)$ was determined with the LCPO method,[73] as implemented in AMBER 7. The surface tension proportionality constant γ and the free energy of nonpolar solvation for a point solute b were set to 0.00542 kcal mol^{-1} Å^{-2} and 0.92 kcal mol^{-1}, respectively, if combined with polar contributions as calculated with Delphi,[63] to 0.0072 kcal mol^{-1} Å^{-2} and 0 kcal mol^{-1}, respectively, in connection with the MGB model,[68] and to 0.005 kcal mol^{-1} and 0 kcal mol^{-1}, respectively, in connection with the vtGB[41] and aoGB[43] models.

Linear relationships of the type of eq. (4) have been widely used in the realm of continuum models to estimate free energies of nonpolar hydration.[35,63,74] Recently, however, some of their deficiencies have been discussed, particularly for macromolecular solvation.[45–49] Considering that nonpolar solvation is determined by the volume and shape of the excluded solvent volume and by attractive van der Waals interactions between solute and solvent, the nonpolar solvation free energy can also be decomposed as[49–51]

$$G^x_{\text{nonpolar}}(i) = G^x_{\text{cav}}(i) + G^x_{\text{vdW,solute–solvent}}(i) \approx G^x_{\text{cav}}(i) + H^x_{\text{vdW,solute–solvent}}(i), \quad (5)$$

where $G^x_{\text{cav}}(i)$ is the cavity solvation free energy and $G^x_{\text{vdW,solute–solvent}}(i)$ is the free energy due to solute–solvent van der Waals dispersion interactions. $G^x_{\text{vdW,solute–solvent}}(i)$ is approximated here by the corresponding van der Waals interaction energy $H^x_{\text{vdW,solute–solvent}}(i)$. Equation (5) thus provides an alternative to eq. (4) for estimating the nonpolar part to solvation free energy.

It has been argued[45,47,50] that the cavity free energy of solvation can be reliably determined by

$$G^x_{\text{cav}}(i) = \gamma_c MSA^x(i), \quad (6)$$

where the surface tension constant γ_c is set to 69 cal mol^{-1} Å^{-2},[75] and $MSA^x(i)$ stands for the molecular surface area. (Recent studies have indicated that this surface definition is superior to the use of the solvent-accessible surface in this case.[75,76]) The molecular surface was determined using the molsurf program of the AMBER 7 suite, applying Bondi radii[64] for the solute atoms and a water probe radius of 1.4 Å. Van der Waals interaction energies $H^x_{\text{vdW,solute–solvent}}(i)$ between solute and solvent atoms were determined using a spherical cutoff of 15 Å. To test the convergence of these values, calculations have been performed for water shells of 7, 8, 9, and 10 Å thickness, respectively, around the solute molecules.

Estimates of Solute Entropy

Entropy contributions arising from changes in the degrees of freedom (translational, rotational, and vibrational) of the solute molecules were included applying classical statistical thermodynamics.[77] On the one hand, contributions to vibrational entropy were obtained by normal-mode analysis. Due to the high computational demand, calculations were performed only for every 10th of the 500 ("set 1," "set 2") or 150 ("set 3") snapshots, respectively, described above. After minimization of each snapshot in the gas phase using the conjugate gradient method with a distance-dependent dielectric of $4r$ (with r being the distance between two atoms) until the root-mean-square of the elements of the gradient vector was less than 10^{-4} kcal mol^{-1} Å^{-1}, frequencies of the vibrational modes were computed at 300 K for these minimized structures including all snapshot atoms and using a harmonic approximation of the energies.

These values are compared to contributions of rotational and vibrational entropy estimated by quasiharmonic analysis,[78] based on the 10-ns production runs, where coordinates had been saved every 100 fs. Because separation of overall rotation from internal motion is not unambiguous for flexible molecules,[79,80] three different ways of least-squares fitting of the snapshots were applied. (1) All atoms of the proteins were used with equal weights. (2) All atoms in helical or strand regions of the proteins were used. (3) Only backbone atoms in helical or strand regions of the proteins were used. In all cases, the snapshot in the middle of the investigated trajectory was taken as the reference structure for superimpositioning.

Table 1. Binding Free Energy Components of the Ras–Raf Complex Calculated for Snapshots Extracted from Separate Trajectories ("set1").[a]

Contribution[c]	Ras–Raf mean[d]	σ[e]	Ras mean[d]	σ[e]	Raf mean[d]	σ[e]	Delta[b] mean[d]	σ[e]
H_{elec}	−10276.5	4.9	−6815.0	4.5	−2233.5	3.0	−1228.0	7.3
H_{vdW}	−1072.4	1.1	−726.4	1.0	−271.6	0.6	−74.4	1.6
H_{int}	4207.8	2.0	2866.9	1.6	1333.0	1.1	7.9	2.8
H_{gas}	−7141.2	5.2	−4674.6	4.7	−1172.2	3.4	−1294.5	7.8
$G_{np,MGB}$	88.3	0.1	62.0	0.1	36.5	0.1	−10.3	0.1
G_{MGB}	−2744.2	4.3	−2528.6	4.2	−1468.5	2.9	1252.9	6.7
$G_{solv,MGB}$	−2655.9	4.3	−2466.6	4.2	−1432.0	2.9	1242.6	6.6
$G_{gas+solv,MGB}$	−9797.1	2.1	−7141.1	1.8	−2604.2	1.2	−51.8	3.0
TS_{total}	2813.1	1.2	1939.4	0.8	908.1	0.6	−34.3	1.6
$G_{total,MGB}$	−12608.4	2.4	−9078.7	2.0	−3510.5	1.3	−19.3	3.4
ΔG_{exp}[f]							−9.7	0.2

[a]All values are given in kcal mol^{-1}. The standard state is taken to be 1 M.
[b]Contribution(Ras–Raf) − contribution(Ras) − contribution(Raf).
[c]H_{elec}: Coulombic energy; H_{vdW}: van der Waals energy; H_{int}: internal energy; $H_{gas} = H_{elec} + H_{vdW} + H_{int}$; $G_{np,MGB}$: nonpolar solvation free energy; G_{MGB}: polar solvation free energy; $G_{solv,MGB} = G_{np,MGB} + G_{MGB}$; $G_{gas+solv,MGB} = H_{gas} + G_{solv,MGB}$; TS_{total}: total entropy contribution as determined by normal mode analysis; $G_{total,MGB} = G_{gas+solv,MGB} + H_{trans/rot} - TS_{total}$.
[d]Average over 500 (respectively 50 in the case of entropy contributions) snapshots.
[e]Standard error of mean values.
[f]Values were determined by isothermal titration calorimetry.[29]

The translational entropy is the only component in the free enthalpy of an ideal solution that depends on solute concentration,[77,81] leading to the concentration dependence of chemical equilibria that do not conserve the number of molecules (such as binding reactions).[82,83] Here, we obtain binding free energies for a standard state of 1 M. It should be noted that this results in a translational entropy for each component that is smaller by 6.4 cal mol^{-1} K^{-1} than the entropy value obtained for the standard state of an (ideal) gas (1 atm, 298.15 K → 0.041 M).

Results and Discussion

Assessment of the Approach

Calculation of absolute binding free energies provides a more stringent test to the underlying methodology used to obtain energy contributions than the pure notion that MD simulations "are stable" (i.e., stay at the correct structure) or evolve to where one would expect the system to go.[16] Now, for example, one can test if the "stable" structure is indeed thermodynamically favorable or just trapped by large barriers. However, cancellation of errors—as advantageously occurs for relative free energy determinations—is lacking, and, thus, even relatively small errors in the individual energetic terms can have a large impact on the (small) final differences.

The MM-PB(GB)SA method used in this study averages contributions of gas-phase energies, solvation free energies, and solute entropies calculated for snapshots of the complex molecule as well as the unbound components, which are extracted from MD trajectories. The binding free energy is then obtained as the difference between the free energy of the complex and the individual molecules. In our case, all but the two water molecules closest to the Mg^{2+} ion of Ras were stripped, as were all counterions. Solvation free energies (including the effects of ionic strengths) are determined by continuum solvation models, which bypasses the problem of adequate sampling over solvent (and counterion) degrees of freedom found in traditional explicit solvent free energy calculations. Finally, compared to traditional approaches for absolute free energy estimates, in the MM-PB(GB)SA approach, no "nonphysical" annihilation[84,85] or decoupling[86,87] of the ligand in solution and bound to its receptor is necessary, nor need one simulate the partially unbound states that would be required for a potential of mean force estimate using umbrella sampling. Hence, binding free energy calculations with MM-PB(GB)SA use physical states at both end points of the binding reaction only, thereby avoiding the need to devote computer time on intermediate states.

General Results

Table 1 contains energetic and entropic contributions to the binding free energy of Ras–Raf, obtained for 500 snapshots (50 in the case of entropy estimates) extracted from the last 10 ns of MD simulations of Raf, Ras, and Ras–Raf ("set 1"). The "gas-phase" energies include Coulomb, van der Waals, and internal energies. The solvation free energy is obtained as sum of a solvent-accessible surface-dependent nonpolar contribution [eq. (4)] and a polar contribution from MGB calculations. Solvation free energy results are reported using a radius of 1.18 Å for the Mg^{2+} ion and salt concentrations of 50 mM ionic strength. The entropic term is obtained as the sum of translational, rotational, and vibrational components, determined by normal mode analysis.

The calculated free energy of Ras–Raf complex formation amounts to -19.3 ± 3.4 kcal mol^{-1}. Compared to the binding free energy of -9.7 ± 0.2 kcal mol^{-1} as determined by isothermal titration calorimetry,[29] the calculation thus overestimates binding. In view of the fact that the accuracy of absolute binding free energy calculations depends on a delicate balance of different energetic and entropic contributions, the calculated binding free energy can still be considered close to the experimentally determined one. One should note, however, that from an experimentalist's point of view, the deviation between experimental and calculated binding free energy obtained here is too large by about one order of magnitude.

We will now turn to the discussion of results in terms of single contributions given in Table 1, but similar findings also hold for Table 4 (see below) and Table A1 (in Supplementary Material). Because the proteins as modeled here carry total charges of 6+ (Raf) and 8− (Ras including GTP and Mg^{2+}), complex formation leads to strongly favorable Coulombic interactions (−1228.0 kcal mol^{-1}), opposed by disfavorable contributions due to the polar part of solvation free energy (1252.9 kcal mol^{-1}). The total electrostatic contribution is 24.9 kcal mol^{-1} and thus disfavors complex formation, which is in general agreement with several other studies (see, e.g., refs. 15 and 18). Interestingly, two studies of the Rap-1A–Raf complex that extract the structures of the unbound proteins from the complex structure (see also below) and allow none[20] or only a limited relaxation[17] of these structures find that electrostatic contributions favor the protein–protein association. It should be noted, however, that a direct comparison to our study is difficult due to the use of different solvation models, ionic strengths of the solutions, and dielectric constants for the solute regions. Contributions from internal energies are also positive (7.9 kcal mol^{-1}) and, hence, disfavor binding. This indicates that conformational changes upon binding lead to internal strain. (In contrast to a related study,[18] our estimated error in the mean value of ΔH_{int} is quite small; thus, the term "structural noise" for this component may not be appropriate here.) Favorable contributions to binding affinity arise from van der Waals interactions (−74.4 kcal mol^{-1}) and the nonpolar part of solvation free energy (−10.3 kcal mol^{-1}), again in agreement with related studies cited above.

Statistical Convergence of Average Energies

To obtain reliable estimates of absolute binding free energies, the average values calculated by MM-PB(GB)SA have to be converged. Figure 1 shows the sum of gas-phase energies and solvation free energies ("effective energies") calculated using the MGB model and a salt concentration of 50 mM as a function of time for snapshots of Raf, Ras, and Ras–Raf ("set 1").

Although these plots indicate rather stable time series of the "effective energies" over the 10 ns of production time of snapshots, the absolute drift still amounts to 4 kcal mol^{-1} ns^{-1} in the case of Ras–Raf, even after 2 ns of unconstrained equilibration. In particular for Raf and Ras, there are long-time oscillations in the "effective energy" and the drifts mainly occur due to incomplete sampling of full cycles. This points to the need of sufficiently long MD trajectories for snapshot extraction. In addition, the "effective energies" significantly fluctuate, hence, mean values of $\Delta G_{\text{gas+solv}}$ can only be estimated to within a standard error of 3.0 kcal mol^{-1} in this case (Table 1). Because the correlation time for decay of fluctuations of the "effective energy" for all systems is about 1 ps, the snapshots used for binding free energy evaluation, which were extracted at time intervals of 20 ps, should be independent. The standard error in the mean values calculated is thus proportional to the inverse of the square root of the number of snapshots.[88] Taken together, the above also implies that the standard error in the mean could be further decreased here by extracting snapshots more frequently from the trajectories.

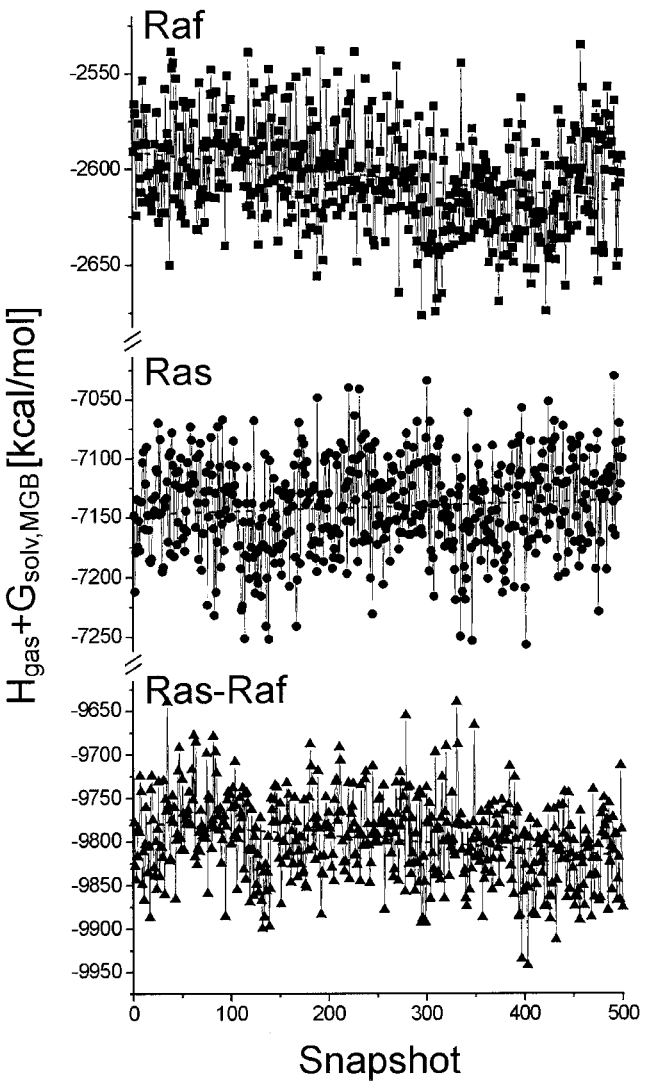

Figure 1. Sum of gas phase and solvation free energies calculated for 500 snapshots extracted at 20-ps intervals from the last 10 ns of MD simulations. The snapshots were taken from separate trajectories of Raf (top), Ras (middle), and Ras–Raf (bottom) ("set 1"). Solvation free energies were calculated using the MGB model at an ionic strength of 50 mM. The slopes of linear regression lines are (values in kcal mol^{-1} ps^{-1}): Raf: $-2*10^{-3}$; Ras: $+1*10^{-3}$; Ras–Raf: $-4*10^{-3}$.

To test the internal consistency of the computations, binding free energy calculations were repeated using only the first or second half of "set 1," respectively (Table 2). A comparison of the

Table 2. Binding Free Energy Components of the Ras–Raf Complex Calculated for the First and Second Half of Snapshots of "set1."[a]

	First half		Second half	
Contribution[b]	Mean[c]	σ[d]	Mean[c]	σ[d]
ΔH_{gas}	−1240.8	9.5	−1348.1	9.5
$\Delta G_{solv,MGB}$	1191.7	7.8	1293.6	8.1
$\Delta G_{gas+solv,MGB}$	−49.1	4.1	−54.5	4.2
$T\Delta S_{total}$	−31.6	2.5	−37.1	1.8
$\Delta G_{total,MGB}$	−19.3	4.8	−19.2	4.6

[a]All values are given in kcal mol^{-1}. The standard state is taken to be 1 M.
[b]Values were calculated according to eq. (1); ΔH_{gas}: gas-phase energy; $\Delta G_{solv,MGB}$: solvation free energy; $\Delta G_{gas+solv,MGB} = \Delta H_{gas} + \Delta G_{solv,MGB}$; $T\Delta S_{total}$: total entropy contribution as determined by normal mode analysis; $\Delta G_{total,MGB} = \Delta G_{gas+solv,MGB} + \Delta H_{trans/rot} - T\Delta S_{total}$.
[c]Average over 250 (respectively 25 in the case of entropy contributions) snapshots.
[d]Standard error of the mean.

results with values obtained for all 500 snapshots (Table 1) indicates essentially identical $\Delta G_{total,MGB}$ values. This agreement results from mutually compensating deviations between values calculated for the full or half data sets for ΔH_{gas}, ΔG_{solv}, and $T\Delta S_{total}$. In that respect, it becomes obvious that accurate estimates of enthalpic or entropic contributions alone are more difficult than free energy estimates.[89] Using only 150 snapshots of the first 3 ns of production runs changes the value of $\Delta G_{total,MGB}$ to −15.0 kcal mol^{-1} (Table A1 in Supplementary Materials), which is still within the error of the calculation. Taking these observations together, the binding free energy components reported thus far are statistically significant, but neither the precision nor the accuracy achieves the level of "chemical accuracy," which is roughly 1 kcal mol^{-1}.

Estimating Solute Entropy

Perhaps the most troublesome component of the binding affinity is the estimate of entropic contributions due to changes in the degrees of freedom of the solutes. Estimates of translational and rotational entropy accompanying association reactions in solution are still a matter of discussion.[90] On the one hand, it has been argued that only the *available* volume within which a molecule is free to move (either in solution or bound in a complex) should be used to estimate translational entropies,[91] leading to the application of concepts of the "cell theory of liquids" to determine these volumes.[92] However, a thoroughly microscopic point of view (which is used here) reveals that the translational (and rotational) entropy change upon binding can be estimated as for two rigid molecules in the gas phase.[81–83] Thus, the *full* volume appropriate to the required standard state is considered. In that respect, it is interesting that a comparison of translational entropies, on the one hand, determined from MD simulations for a β-heptapeptide in explicit solvent and, on the other, by applying ideal gas results agreed very well.[93]

In the present study, we applied classical statistical thermodynamics. As in previous applications of MM-PB(GB)SA,[35,41] on the one hand, entropic contributions have been calculated using normal mode analysis, which approximates the potential energy surface of a molecular system by a harmonic function at some point of the surface. Hence, transitions from one local minimum to another are neglected.[94] Contributions due to the loss of translational and rotational degrees of freedom (TS_{trans} and TS_{rot}) disfavor complex formation by about 14 and 15 kcal mol^{-1}, respectively (Table 3). In addition, changes in vibrational degrees of freedom as determined by normal mode analysis[94] amount to $TS_{vib} = -5.8$ kcal mol^{-1}. Estimates of errors in the mean values of $TS_{total,NM}$ as given in Table 1 almost completely result from values obtained for $TS_{vib,NM}$. In this respect, an error in the mean value of $TS_{vib,NM}$ of 1.6 kcal mol^{-1} (determined from fluctuations across 50 different snapshots) indicates that structures at different local energy minima rather give very similar vibrational entropy estimates. Although the rotational entropic component depends on the molecule's principal moments of inertia and, hence, the conformation of the molecule, variations in $TS_{rot,NM}$ as computed for different snapshots of "set 1" were found to be vanishing.

An alternative is given by quasiharmonic analysis, where effective modes are computed such that the second moments of the amplitude distribution match those found in the (explicit solvent) MD simulation.[78,95] Consequently, some of the effects of anharmonic terms in the force field are included in the analysis. Essentially, identical rotational entropy contributions were obtained by the quasiharmonic analyses compared to normal-mode analyses (Table 3). As Table 3 indicates, however, the vibrational entropy components strongly deviate from the results found by normal-mode analyses, which leads to very disfavorable (and unreasonable) overall entropic contributions. The results from three different schemes applied to superimpose the snapshots deviate by 4.7 kcal/mol, indicating that the way overall rotational motion is removed has a significant influence on the results.

Table 3. Entropic Contributions of the Ras–Raf System Estimated by Normal Mode and Quasiharmonic Analyses.[a]

Contribution[b]	Ras–Raf	Ras	Raf	Delta[c]
TS_{trans}	15.0	14.7	14.0	−13.7
$TS_{rot,NM}$	17.4	16.7	15.6	−14.9
$TS_{vib,NM}$	2780.6	1907.9	878.5	−5.8
$TS_{total,NM}$	2813.1	1939.4	908.1	−34.3
$TS_{rot,QH}$	17.5	16.8	15.6	−14.9
$TS_{vib,QH1}$	3887.2	2695.9	1431.7	−240.4
$TS_{vib,QH2}$	3888.4	2697.4	1433.9	−242.8
$TS_{vib,QH3}$	3890.1	2699.1	1436.1	−245.1
$TS_{total,QH1}$	3921.7	2729.3	1463.2	−270.9
$TS_{total,QH2}$	3922.9	2730.8	1465.4	−273.3
$TS_{total,QH3}$	3924.6	2732.5	1467.6	−275.6

[a]All values are given in kcal mol^{-1}. The standard state is taken to be 1 M. Calculations were performed for every 10th snapshot of "set 1" in the case of normal mode analyses and for snapshots saved every 100 fs along the 10 ns long trajectories for quasiharmonic analyses.
[b]$TS_{trans/rot/vib/total}$: translational/rotational/vibrational/total entropy; NM: normal mode analysis; QH1/2/3: quasiharmonic analysis using least-squares fitting schemes 1, 2, or 3, respectively.
[c]Contribution(Ras–Raf) − contribution(Ras) − contribution(Raf).

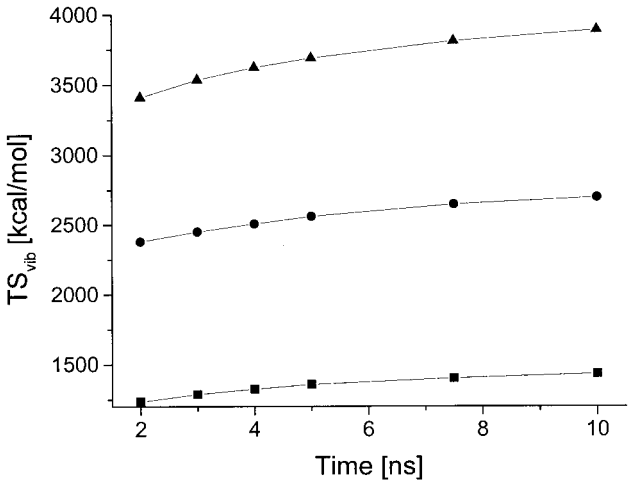

Figure 2. Vibrational entropy contributions TS_{vib} at $T = 300$ K as computed by quasiharmonic analyses for different lengths of the MD trajectories of Raf (■), Ras (●), and Ras–Raf (▲). Snapshots were sampled every 100 fs, and least-squares fitting was performed including all atoms of the molecules. The lines were obtained by connecting the data points.

Figure 2 shows results of vibrational entropies computed by quasiharmonic analyses from segments of the MD trajectories of 2, 3, 4, 5, 7.5, and 10-ns length. None of the three proteins yield convergent vibrational entropy contributions within 10 ns of simulation time. An attempt to extrapolate the curves to infinite time yielded large statistical uncertainties (data not shown). In a recent study[96] on human α-lactalbumin using MD simulations of 2 ns length, calculations of the configurational entropy for the entire protein and for the backbone and sidechains did not converge either. Along these lines, configurational entropy estimates for a β-heptapeptide even did not show convergence after 150 ns of simulation time.[93,97] In total, although quasiharmonic analysis provides an interesting alternative to normal-mode analysis for estimating absolute entropies of solute molecules, sampling issues may pose a limitation to its straightforward applicability to problems like this one.

"Single-Trajectory" Calculations vs. "Three-Trajectories" Calculations

In Table 4, we report contributions to the binding free energies of Ras–Raf calculated according to eq. (3) for snapshots extracted from the single complex trajectory ("set 2"). The method is much cheaper than the one using snapshots from three separate trajectories and potentially may require less sampling, because all of the intramolecular energies cancel when calculating the association energy.

In related studies applying MM-PB(GB)SA for the prediction of absolute binding affinities,[15,41] good agreement between values calculated for snapshots extracted either from the complex trajectory or from all three trajectories has been found. This finding has then been interpreted in terms of the well-defined and rather rigid binding interfaces of the systems investigated. Similarly, studies extracting unbound molecules from only the crystal structure of the complex and allowing none or restricted relaxation by minimization have been reported.[17,19,20] This approach is only successful, however, if no conformational and dynamical changes upon complex formation occur in the binding partners.[18,19] In the case of Ras, there is evidence from ^{31}P NMR experiments that the unbound molecule exists in at least two conformational states and that one of the conformations is stabilized upon effector binding.[98,99] Furthermore, minor changes in the switch region drastically affect dynamic behavior of Ras and, in turn, interaction with effectors.[100] In that sense, it is not surprising that binding free energy calculations in the present study yielded different results for ΔH_{gas}, $\Delta G_{solv,MGB}$, and $T\Delta S_{total}$ if using snapshots of "set 1" or "set 2" (Tables 1 and 4).

Table 4 reveals that in total the binding affinity of Raf towards Ras is underestimated ($\Delta G_{total,MGB} = -3.1$ kcal mol^{-1}) when

Table 4. Binding Free Energy Components of the Ras–Raf Complex Calculated for Snapshots Extracted from the Complex Trajectory ("set2").[a]

	Ras–Raf		Ras		Raf		Delta	
Contribution	Mean	σ	Mean	σ	Mean	σ	Mean	σ
H_{elec}	−10276.5	4.9	−6913.2	3.1	−2115.7	2.7	−1247.7	2.5
H_{vdW}	−1072.4	1.1	−734.6	1.0	−276.9	0.6	−60.9	0.2
H_{int}	4207.8	2.0	2872.0	1.7	1335.8	1.1	0.0	0.0
H_{gas}	−7141.2	5.2	−4775.8	3.2	−1056.8	2.9	−1308.6	2.5
$G_{np,MGB}$	88.3	0.1	61.0	0.1	36.7	0.1	−9.5	0.0
G_{MGB}	−2744.2	4.3	−2444.0	2.7	−1575.6	2.3	1275.3	2.4
$G_{solv,MGB}$	−2655.9	4.3	−2382.9	2.7	−1538.9	2.2	1265.9	2.4
$G_{gas+solv,MGB}$	−9797.1	2.1	−7158.7	1.6	−2595.7	1.2	−42.7	0.3
TS_{total}	2813.1	1.2	1942.3	0.9	912.2	0.6	−41.4	1.6
$G_{total,MGB}$	−12608.4	2.4	−9099.2	1.9	−3506.1	1.4	−3.1	1.6
ΔG_{exp}							−9.7	0.2

[a]See legend of Table 1; however, "delta" values were calculated according to eq. (3).

snapshots of "set 2" are used. While the gas-phase energy $\Delta H_{gas} = -1308.6$ kcal mol^{-1} is even more favorable here compared to the result in Table 1, the desolvation penalty ($\Delta G_{solv,MGB} = 1265.9$ kcal mol^{-1}) and the entropic contributions due to changes in the degrees of freedom of the solutes ($T\Delta S_{total} = -41.4$ kcal mol^{-1}) become less favorable than in the case of snapshots extracted from separate trajectories. Comparing contributions of internal energy for Raf and Ras either in the unbound state (Table 1) or extracted from the respective complex trajectory (Table 4) reveals that upon binding Ras and Raf experience conformational strain as would be expected.

Due to the adaptation of binding partners to each other upon association, the free energy of the bound conformation of a molecular species is expected to be higher than the one obtained for the unbound conformation. (If this were not the case, the isolated molecule would predominantly exist in the bound conformation in solution.) Thus, $G_{total,MGB}$ values of Raf or Ras should be lower for snapshots extracted from separate trajectories (representing the unbound conformation) than for snapshots extracted from the complex trajectory (representing the bound conformation), yielding a less favorable complex formation in the "three-trajectories" case.

Here, however, we find the opposite: the binding affinity calculated from snapshots extracted from the separate trajectories ($\Delta G_{total,MGB} = -19.3$ kcal mol^{-1}) is more negative than the one obtained for snapshots extracted from the complex trajectory. Raf snapshots show a lower free energy of stability in the unbound conformation (separate trajectory: -3510.5 kcal mol^{-1}) compared to the bound conformation (complex trajectory: -3506.1 kcal mol^{-1}), as expected. However, for Ras, $G_{total,MGB}$ values favor snapshots from the complex trajectory by ca. 20 kcal mol^{-1}. (This also holds if only $G_{gas+solv,MGB}$ values are considered.) The finding may raise the question whether the ensemble of structures obtained by the explicit solvent simulations is adequately represented by the results of implicit solvent free energy calculations. This suggests that the implicit solvent model imperfectly reflects the forces experienced in the explicit solvent simulation and underlines the difficulties of reliably estimating strain energies of macromolecules.

Comparison of Generalized Born and Poisson–Boltzmann Results

Upon Ras–Raf complex formation, polar and charged residues become buried in the binding interface. Modeling of the desolvation penalty and correctly estimating screened electrostatic interactions between both proteins is thus crucial for the accurate estimation of binding affinity.

To test the influence of the results of different continuum solvation models on the absolute binding free energy of Ras–Raf, solvation free energy contributions were calculated by solving the Poisson–Boltzmann equation using radii from the PARSE parameter set[63] (linearized PB: PB1, nonlinearized PB: PB2) or using Bondi radii[64] (linearized PB: PB3), or applying generalized Born models of Jayaram et al.[68] (MGB), Onufriev et al.[43] (aoGB), and Tsui et al.[41] (vtGB). All GB models are based on the analytical method of Hawkins et al.[71] for the calculation of effective Born radii, with a modification for aoGB that was recently proposed to permit a more accurate description of large macromolecules. MGB uses a modification of the standard[65] form of $f_{ij}^{GB}(r_{ij})$, which was introduced to obtain a better description of intramolecular interactions. Due to the computational demand, the following calculations will be reported for 150 snapshots extracted from the first 3 ns of production time of the separate trajectories ("set 3"). As described above, using only a subset of snapshots did not lead to a significant change of the calculated binding free energy compared to the one obtained for the total "set 1." Table A1 (in Supplementary Material) summarizes the contributions to the binding free energy. In all continuum solvation model calculations, atomic radii and parameters for the nonpolar part of the solvation free energy were used as reported in the respective studies. Hence, absolute values of different models may not be compared to each other.

Calculated absolute binding free energies ΔG_{total} strongly depend on the applied solvation model. Although within the continuum solvent approximation solvation free energy computations by solving the Poisson(–Boltzmann) equation are often considered to be a target for GB optimization, in the present study, binding free energy calculations applying PB calculations resulted in considerable over- or underbinding of both proteins, depending on the atomic radii applied ($\Delta G_{total,PB1} = 5.6$ kcal mol^{-1}, $\Delta G_{total,PB3} = -34.9$ kcal mol^{-1}, Table A1 in Supplementary Material). The observed radii dependency can be explained by the fact that PARSE radii are, on average, smaller than Bondi radii, placing the dielectric boundary closer to atomic sites in the former case. This, in turn, leads to lower energies for polar atoms because of stronger polarization in the case of PARSE radii, compared to Bondi radii, which is expected to be more pronounced for exposed atoms than for buried ones. Taking into account that upon complex formation, polar atoms become (partially) buried in the Ras–Raf interface, a higher desolvation penalty (and, hence, less favorable binding) can be expected for PB calculations applying PARSE radii. Repeating the calculations using the nonlinearized PB equation together with PARSE radii resulted in $\Delta G_{total,PB2} = 6.2$ kcal mol^{-1}, which is very similar to the result obtained with the linearized PB equation (PB1).

Generalized Born methods can be considered as an approximation to the computationally intensive PB approach. Application of three different GB models results in an overestimation of the binding free energies of Ras–Raf in general, with the MGB model yielding the best agreement with the experimental affinity (exp.: -9.7 kcal mol^{-1}; calc.: -15.0 kcal mol^{-1}). Yet, the calculated solvation free energy difference $\Delta G_{solv,MGB}$ differs considerably from the ones obtained by PB. Application of the aoGB and vtGB models lead to even larger deviations with respect to experimental and theoretical binding affinities ($\Delta G_{total,aoGB} = -32.6$ kcal mol^{-1}; $\Delta G_{total,vtGB} = -49.4$ kcal mol^{-1}). In all cases, the nonpolar contributions to solvation free energies are within 1.4 kcal mol^{-1} and the differences thus arise from the electrostatic part of the calculations.

Figure 3 compares electrostatic contributions to the solvation free energy of Raf, Ras, and Ras–Raf as calculated by the GB models (vtGB,[41] MGB,[68] and aoGB[43]) with those obtained by solving the Poisson equation (PE), using the same van der Waals radii and charges in the PE calculations as in the GB calculations. The scatterplots reveal that in all cases GB results deviate systematically from PB results. Table 5 summarizes statistical parameters

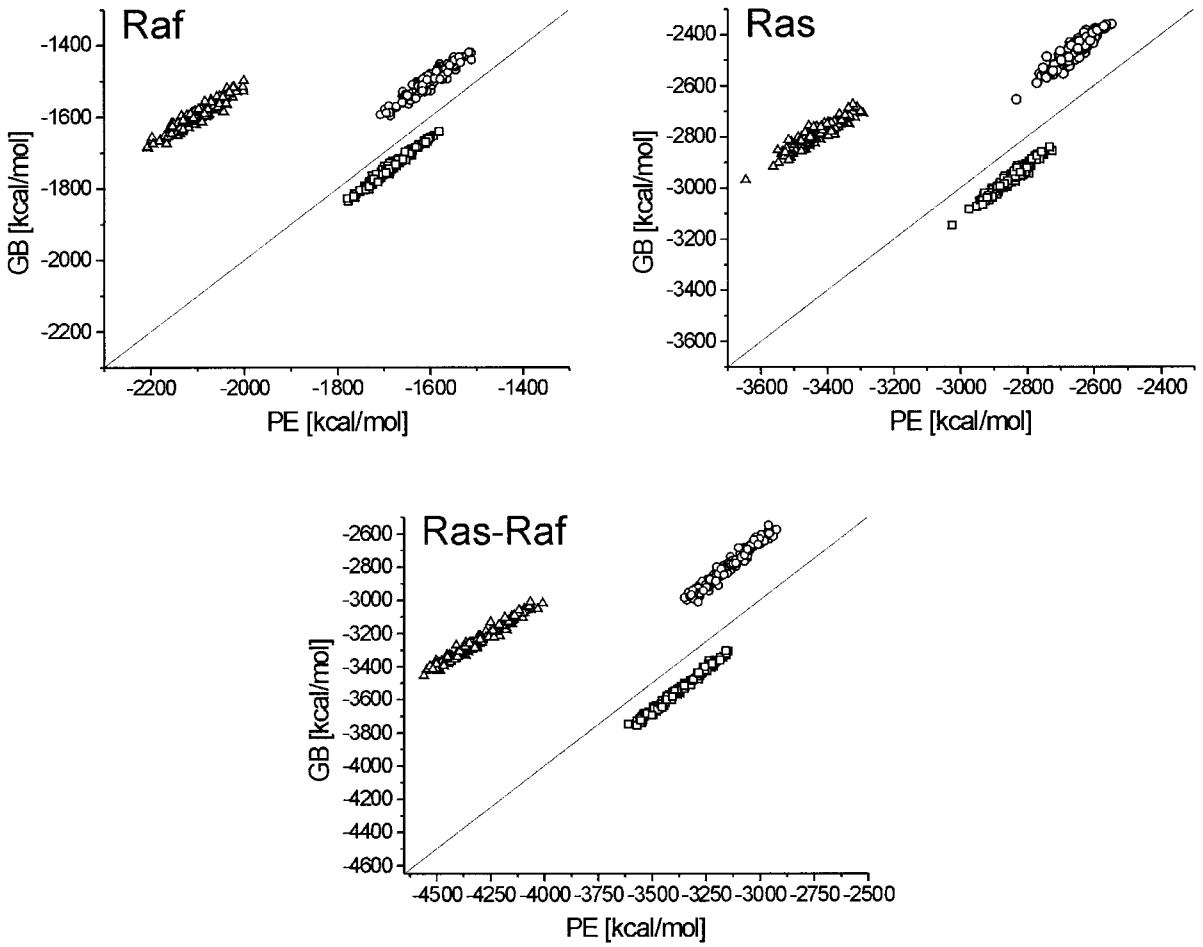

Figure 3. Comparison of the polar part of solvation free energy as calculated by solving the Poisson equation (PE) or applying the vtGB (△), aoGB (□), and MGB (○) models for 150 snapshots of Raf (top left), Ras (top right), and Ras–Raf (bottom), extracted at 20-ps intervals from 3 ns of MD simulation after 2 ns of equilibration ("set 3"). In all cases, corresponding radii were used for PE and GB calculations. The line represents a perfect match between the PE and GB values.

of the correlations. The correlation coefficient and root-mean-square deviation reflect the ability of the GB models to score different snapshots of the same molecule in the same way as the corresponding PE calculations. All of the GB models are reasonably successful at this task, with correlation coefficients above 0.93 in all cases. The aoGB model gives the best agreement

Table 5. Statistical Parameters of the Comparison of Electrostatic Solvation Free Energies Calculated Using Generalized Born Models vs. Poisson Calculations.

	Ras–Raf			Ras			Raf		
GB model	R[a]	Rmsd[b]	ME[c]	R[a]	Rmsd[b]	ME[c]	R[a]	Rmsd[b]	ME[c]
MGB	0.978	22.9	350.2	0.931	19.8	193.7	0.951	12.6	102.2
aoGB	0.993	13.2	−166.9	0.972	12.12	−114.7	0.987	6.5	−55.6
vtGB	0.983	31.7	1061.9	0.935	24.0	640.4	0.952	14.6	504.5

[a]Correlation coefficient.
[b]Root-mean-square deviation calculated as $\langle(\Delta G_{GB,i} - \Delta G_{PB,i})^2\rangle_i^{1/2}$ (where $\Delta G_{X,i} = G_{X,i} - \langle G_{X,j}\rangle_j$) and given in kcal mol^{-1}. $\langle \cdots \rangle_i$ indicates an average over all snapshots i.
[c]Mean error calculated as $\langle G_{GB,i}\rangle_i - \langle G_{PB,i}\rangle_i$ and given in kcal mol^{-1}.

between PE and GB calculations in terms of relative solvation free energies for different conformations of Raf, Ras, and Ras–Raf. This is not unexpected, because aoGB has been modified to give better agreement between PE and GB results for molecules with larger interior regions.[43] Where these models are less successful, however, is in reproducing the Poisson results for the absolute solvation energies of different proteins. This is reflected in the "ME" column, which shows significant differences in the mean energies computed by GB and Poisson theories for all three molecular systems.

The above findings, in particular for aoGB and vtGB, may result from the fact that these models have been parameterized primarily for the use in molecular dynamics simulations, i.e., they have been optimized to reproduce energy differences between different conformations of one molecule in agreement with PE results. For computation of binding free energies, however, parameter sets will have to be developed, which also reproduce solvation free energies across different (macro-)molecular species.

Influence of the Nonpolar Solvation Contribution

So far, a solvent-accessible surface area-dependent model has been applied to estimate the total nonpolar solvation free energy [eq. (4)]. Recent studies have discussed deficiencies of these surface area models, in particular in the case of biomolecular solvation. It has been shown that favorable van der Waals attraction between interior atoms of the solute and solvent atoms may not be captured[45,47,48] and this effect has been estimated to account for up to 40% of the total van der Waals protein–water interaction energy.[49] Consequently, favorable van der Waals dispersion between solute and solvent atoms needs to be considered, especially when the number of solvent-exposed and buried atoms differs considerably, as in protein–protein complex formation. Applying a small molecule-derived surface area model instead omits favorable interactions of the unbound molecules with the solvent and, thus, overestimates the contribution of nonpolar desolvation to the binding free energy.

By estimating the nonpolar solvation contribution as sum of a surface area-dependent cavity term and explicitly modeled van der Waals interactions between solute and solvent atoms [eq. (5)], these favorable contributions due to buried atoms can be included. Approximating these free energy contributions by van der Waals interaction energies may be justified by findings from energy decomposition in that the process of adding solute–solvent interactions after cavity formation is nearly totally enthalpic,[47] i.e., these interactions do not affect the solvent structure. One should note, however, that, in contrast to a recent study applying the attractive part of the van der Waals potential only,[49] eq. (5) includes both dispersion and repulsion contributions. In this respect, our approach is similar in spirit to the ES/IS model of Vorobjev et al.[101]

Table 6 shows solute-explicit solvent van der Waals interaction energies averaged over 500 snapshots of Raf, Ras, and Ras–Raf, respectively, extracted from separate trajectories. Repeating the calculations for water shells of 7, 8, 9, and 10 Å thickness shows that the $\Delta H_{\mathrm{vdW,solute-solvent}}$ values obtained for shells of 9 and 10 Å thickness differ by amounts within the standard errors, indicating a convergence of the results with respect to the size of the water shell considered. Values determined only over the first or second half of the snapshots, respectively, are also within the standard error in the mean (data not shown).

Table 6. Solute-Explicit Solvent van der Waals Interaction Energies.[a]

Shell[b]	Ras–Raf	Ras	Raf	Delta[c]
7	−388.6	−249.2	−186.2	46.8
8	−398.9	−257.4	−191.0	49.5
9	−404.7	−262.1	−193.8	51.2
10	−407.8	−264.7	−195.4	52.3
σ[d]	1.4	1.2	0.8	2.0

[a]All values are given in kcal mol^{-1}. The values were averaged over 500 snapshots of Raf, Ras, and Ras–Raf, respectively, extracted from separate trajectories ("set 1").
[b]Thickness (in Å) of the solvation shell around the solute.
[c]Contribution(Ras–Raf) − contribution(Ras) − contribution(Raf).
[d]Standard error of mean values. These values are identical to the given precision for all shells considered.

The difference in solute–solvent van der Waals interaction energies determined for the 10 Å water shell amounts to $\Delta H_{\mathrm{vdW,solute-solvent}} = 52.3$ kcal mol^{-1}. Because the solute–solute van der Waals energy is −60.9 kcal mol^{-1} (see ΔH_{vdW} in Table 4), the overall contribution to the binding affinity due to changes in the dispersion-repulsion interactions is favorable. The cavity free energy calculations [eq. (6)] yield G_{cav} average values for Raf, Ras, and Ras–Raf of 306.9 ± 0.4, 544.9 ± 0.4, and 800.0 ± 0.5 kcal mol^{-1}, respectively, resulting in a contribution to the binding free energy of $\Delta G_{\mathrm{cav}} = -51.7 \pm 0.8$ kcal mol^{-1}. Together with $\Delta H_{\mathrm{vdW,solute-solvent}} = 52.3$ kcal mol^{-1}, the total nonpolar solvation component determined with eq. (5) then amounts to $\Delta G_{\mathrm{np}} = 0.6 \pm 2.2$ kcal mol^{-1}. As anticipated, this contribution is less favorable than the one calculated with the solvent-accessible surface area-dependent model [eq. (4); −10.3 kcal mol^{-1}; Table 1).

Adding the gas-phase contributions, the polar part of solvation free energy, and the entropic contributions due to changes in the degrees of freedom of the binding partners from Table 1 to ΔG_{np} finally yields a binding free energy $\Delta G_{\mathrm{total,GB}}$ of −8.3 ± 3.2 kcal mol^{-1} for Ras–Raf. This value is in good agreement with the experimentally determined binding free energy of −9.6 kcal mol^{-1}. In view of the dependency of the result on the choice of the continuum model used to estimate the polar part of solvation free energy, however, testing the transferability of the approach that led to this result on an increased data set is necessary.

Conclusion

In the present study, absolute binding free energies have been calculated using the MM-PB(GB)SA approach for the protein–protein complex Ras–Raf, i.e., gas-phase and solvation energies complemented by entropic estimates for the solute molecules have been averaged over snapshots extracted from MD simulations. The internal consistency and the model dependence of the results have been tested. The calculated binding free energy values should be

reproducible from independent trajectories. The results obtained using different GB models in connection with a solvent-accessible surface area-dependent term to estimate nonpolar solvation overestimate the binding affinity of Raf towards Ras, whereas PB calculations yield binding free energies that are either too positive or too negative, depending on the atomic radii applied. Calculating the nonpolar solvation contribution by a sum of a cavity solvation free energy and explicitly modeled solute–solvent van der Waals interactions yields less favorable contributions to the binding free energy compared to the solvent-accessible surface area model. Entropy estimates by normal mode analyses indicate that structures at different local energy minima provide rather similar rotational and vibrational contributions. Applying quasiharmonic analyses, however, no convergence of the vibrational entropy estimates within 10 ns of simulation time could be observed.

Binding free energies have also been computed for snapshots of the unbound molecules extracted from the complex trajectory. Here, use of the MGB model results in an underestimation of the binding affinity, compared to experimental results. In view of the conformational and dynamical changes experimentally determined for Ras upon binding, this deviation with respect to the results obtained for snapshots from three separate trajectories is not astonishing. Because these changes in many cases cannot be observed on the time scales available to standard MD simulations today, one has to exercise caution even if starting structures for unbound (separate) simulations are extracted from an experimental complex structure. Furthermore, while the influence of molecular strain experienced by the protein molecules upon binding is correctly estimated for Raf, Ras snapshots from the complex trajectory are more favorable than Ras snapshots extracted from the independent trajectory. This finding has not yet been understood.

The influence of mobile water molecules even in the vicinity of polar or charged groups should be correctly described by the continuum solvent models, yet, it is arguable whether highly restricted water molecules should be considered as part of the continuum or whether local interactions modulated by "structural water molecules" should be considered by an explicit description. Along these lines, one water molecule has been identified in the binding interface of Ras–Raf that shows atomic fluctuations below 1 Å and mediates interactions between both binding partners. In this study, we refrained from including the "structural water molecule" in our calculations. However, in view of a contribution to binding affinity of -1.6 kcal mol^{-1} found for a solvent-mediated interaction by assuming ideal interaction geometries,[102] an adequate handling of "structural water molecules" will become mandatory.

The current study underlines that absolute binding free energy calculations are still very demanding, especially if performed for macromolecular systems where a rather large amount of surface area is buried upon complex formation and, hence, the region of lower dielectric expands considerably. Although the best estimate for the binding free energy of Ras–Raf obtained in this study of -8.3 kcal mol^{-1} is in good agreement with the experimental result of -9.6 kcal mol^{-1}, further testing and parameterizations across large test sets of molecules with different properties will be necessary to probe the transferability of the protocols and further improve parameters. In this respect, it will be interesting to evaluate the influence of recent developments in the field of polarizable force fields and continuum solvent descriptions on the outcome of these calculations. Nevertheless, even at this early stage, the combination of MD simulations with a subsequent analysis of energetic and entropic components provides an exciting opportunity to obtain increased insights into the determinants of molecular association.

Acknowledgments

We thank Dr. A. Onufriev (The Scripps Research Institute, La Jolla) for helpful discussions. H.G. gratefully acknowledges a Feodor-Lynen fellowship awarded by the Alexander-von-Humboldt foundation, Germany.

References

1. Stites, W. E. Chem Rev 1997, 97, 1233.
2. Kleanthous, C., Ed. Protein–Protein Recognition; Oxford University Press: Oxford, 2000, vol. 31.
3. Valencia, A.; Pazos, F. Curr Opin Struct Biol 2002, 12, 368.
4. Teichmann, S. A.; Murzin, A. G.; Chothia, C. Curr Opin Struct Biol 2001, 11, 354.
5. Jones, S.; Thornton, J. M. Proc Natl Acad Sci USA 1996, 93, 13.
6. Davies, D. R.; Cohen, G. H. Proc Natl Acad Sci USA 1996, 93, 7.
7. Brooijmans, N.; Sharp, K. A.; Kuntz, I. D. Proteins 2002, 48, 645.
8. Elcock, A. H.; Sept, D.; McCammon, J. A. J Phys Chem B 2001, 105, 1504.
9. Novotny, J.; Bruccoleri, R. E.; Saul, F. A. Biochemistry 1989, 28, 4735.
10. Guerois, R.; Nielsen, J. E.; Serrano, L. J Mol Biol 2002, 320, 369.
11. Kortemme, T.; Baker, D. Proc Natl Acad Sci USA 2002, 99, 14116.
12. Moont, G.; Gabb, H. A.; Sternberg, M. J. Proteins 1999, 35, 364.
13. Jiang, L.; Gao, Y.; Mao, F.; Liu, Z.; Lai, L. Proteins 2002, 46, 190.
14. Brandsdal, B. O.; Aqvist, J.; Smalas, A. O. Protein Sci 2001, 10, 1584.
15. Massova, I.; Kollman, P. A. J Am Chem Soc 1999, 121, 8133.
16. Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.; Cieplak, P.; Srinivasan, J.; Case, D. A.; Cheatham, T. E., III. Acc Chem Res 2000, 33, 889.
17. Muegge, I.; Schweins, T.; Warshel, A. Proteins 1998, 30, 407.
18. Huo, S.; Massova, I.; Kollman, P. A. J Comput Chem 2002, 23, 15.
19. Noskov, S. Y.; Lim, C. Biophys J 2001, 81, 737.
20. Sheinerman, F. B.; Honig, B. J Mol Biol 2002, 318, 161.
21. Wittinghofer, A.; Waldmann, H. Angew Chem Int Ed Engl 2000, 39, 4193.
22. Wittinghofer, A.; Nassar, N. Trends Biochem Sci 1996, 21, 488.
23. Emerson, S. D.; Madison, V. S.; Palermo, R. E.; Waugh, D. S.; Scheffler, J. E.; Tsao, K. L.; Kiefer, S. E.; Liu, S. P.; Fry, D. C. Biochemistry 1995, 34, 6911.
24. Pai, E. F.; Kabsch, W.; Krengel, U.; Holmes, K. C.; John, J.; Wittinghofer, A. Nature 1989, 341, 209.
25. Pai, E. F.; Krengel, U.; Petsko, G. A.; Goody, R. S.; Kabsch, W.; Wittinghofer, A. EMBO J 1990, 9, 2351.
26. Milburn, M. V.; Tong, L.; deVos, A. M.; Brunger, A.; Yamaizumi, Z.; Nishimura, S.; Kim, S. H. Science 1990, 247, 939.
27. Nassar, N.; Horn, G.; Herrmann, C.; Scherer, A.; McCormick, F.; Wittinghofer, A. Nature 1995, 375, 554.

28. Nassar, N.; Horn, G.; Herrmann, C.; Block, C.; Janknecht, R.; Wittinghofer, A. Nat Struct Biol 1996, 3, 723.
29. Rudolph, M. G.; Linnemann, T.; Grunewald, P.; Wittinghofer, A.; Vetter, I. R.; Herrmann, C. J Biol Chem 2001, 276, 23914.
30. Chen, J. M.; Monaco, R.; Manolatos, S.; Brandt–Rauf, P. W.; Friedman, F. K.; Pincus, M. R. J Protein Chem 1997, 16, 619.
31. Zeng, J.; Treutlein, H. R.; Simonson, T. Proteins 1999, 35, 89.
32. Zeng, J.; Treutlein, H. R.; Simonson, T. Proteins 1998, 31, 186.
33. Zeng, J.; Fridman, M.; Maruta, H.; Treutlein, H. R.; Simonson, T. Protein Sci 1999, 8, 50.
34. Gohlke, H.; Kiel, C.; Case, D. A. J Mol Biol 2003, 330, 891.
35. Srinivasan, J.; Cheatham, T. E., III; Cieplak, P.; Kollman, P. A.; Case, D. A. J Am Chem Soc 1998, 120, 9401.
36. Lee, M. R.; Baker, D.; Kollman, P. A. J Am Chem Soc 2001, 123, 1040.
37. Kuhn, B.; Kollman, P. A. J Am Chem Soc 2000, 122, 3909.
38. Lee, T.; Kollman, P. A. J Am Chem Soc 2000, 122, 4385.
39. Wang, J.; Morin, P.; Wang, W.; Kollman, P. A. J Am Chem Soc 2001, 123, 5221.
40. Wang, W.; Kollman, P. A. J Mol Biol 2000, 303, 567.
41. Tsui, V.; Case, D. A. J Phys Chem B 2001, 105, 11314.
42. Massova, I.; Kollman, P. A. Perspect Drug Discov Design 2000, 18, 113.
43. Onufriev, A.; Bashford, D.; Case, D. A. J Phys Chem B 2000, 104, 3712.
44. Lee, M. S.; Salsbury, F. R.; Brooks, C. L. J Chem Phys 2002, 116, 10606.
45. Ashbaugh, H. S.; Kaler, E. W.; Paulaitis, M. E. J Am Chem Soc 1999, 121, 9243.
46. Lum, K.; Chandler, D.; Weeks, J. D. J Phys Chem B 1999, 103, 4570.
47. Gallicchio, E.; Kubo, M. M.; Levy, R. M. J Phys Chem B 2000, 104, 6271.
48. Pitera, J. W.; van Gunsteren, W. F. J Am Chem Soc 2001, 123, 3163.
49. Levy, R. M.; Zhang, L. Y.; Gallicchio, E.; Felts, A. K. J Am Chem Soc 2003, 125, 9523.
50. Tunon, I.; Silla, E.; Pascual–Ahuir, J. L. Chem Phys Lett 1993, 203, 289.
51. Pratt, L. R.; Chandler, D. J Chem Phys 1977, 67, 3683.
52. Case, D. A.; Pearlman, D. A.; Caldwell, J. W.; Cheatham, T. E., III; Wang, J.; Ross, W. S.; Simmerling, C. L.; Darden, T. A.; Merz, K. M.; Stanton, R. V.; Cheng, A. L.; Vincent, J. J.; Crowley, M.; Tsui, V.; Gohlke, H.; Radmer, R. J.; Duan, Y.; Pitera, J.; Massova, I.; Seibel, G. L.; Singh, U. C.; Weiner, P. K.; Kollman, P. A.; University of California: San Francisco, 2002.
53. Cornell, W. D.; Cieplak, C. I.; Bayly, I. R.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. J Am Chem Soc 1995, 117, 5179.
54. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J.; Klein, M. L. J Chem Phys 1983, 79, 926.
55. Leach, A. R.; Klein, T. E. J Comput Chem 1995, 16, 1378.
56. Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. J Phys Chem 1993, 97, 10269.
57. Aqvist, J. J Mol Struct (Theochem) 1992, 88, 135.
58. Darden, T.; York, D.; Pedersen, L. J Chem Phys 1993, 98, 10089.
59. Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. J Comput Phys 1977, 23, 327.
60. Bernstein, F. C.; Koetzle, T. F.; Williams, G. J.; Meyer, E. E., Jr.; Brice, M. D.; Rodgers, J. R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. J Mol Biol 1977, 112, 535.
61. Terada, T.; Ito, Y.; Shirouzu, M.; Tateno, M.; Hashimoto, K.; Kigawa, T.; Ebisuzaki, T.; Takio, K.; Shibata, T.; Yokoyama, S.; Smith, B. O.; Laue, E. D.; Cooper, J. A. J Mol Biol 1999, 286, 219.
62. DelPhi/Solvation; Molecular Simulations Inc.: San Diego, CA, 1995.
63. Sitkoff, D.; Sharp, K. A.; Honig, B. J Phys Chem 1994, 98, 1978.
64. Bondi, A. J Chem Phys 1964, 64, 441.
65. Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. J Am Chem Soc 1990, 112, 6127.
66. Schaefer, M.; Froemmel, C. J Mol Biol 1990, 216, 1045.
67. Bashford, D.; Case, D. A. Annu Rev Phys Chem 2000, 51, 129.
68. Jayaram, B.; Sprous, D.; Beveridge, D. L. J Phys Chem B 1998, 102, 9571.
69. Minehardt, T. J.; Cooke, R.; Pate, E.; Kollman, P. A. Biophys J 2001, 80, 1151.
70. Badenhoop, J. K.; Weinhold, F. J Chem Phys 1997, 107, 5422.
71. Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. J Phys Chem 1996, 100, 19824.
72. Srinivasan, J.; Trevathan, M. W.; Beroza, P.; Case, D. A. Theor Chem Acc 1999, 101, 426.
73. Weiser, J.; Shenkin, P. S.; Still, W. C. J Comput Chem 1999, 20, 217.
74. Lee, M. R.; Duan, Y.; Kollman, P. A. Proteins 2000, 39, 309.
75. Jackson, R. M.; Sternberg, M. J. E. Protein Eng 1994, 7, 371.
76. Pitarch, J.; Moliner, V.; Pascual–Ahuir, J. L.; Silla, E.; Tunon, I. J Phys Chem 1996, 100, 9955.
77. McQuarrie, D. A. Statistical Mechanics; Harper & Row: New York, 1976.
78. Teeter, M. M.; Case, D. A. J Phys Chem 1990, 94, 8091.
79. Hunenberger, P. H.; Mark, A. E.; van Gunsteren, W. F. J Mol Biol 1995, 252, 492.
80. Karplus, M.; Ichiye, T. J Mol Biol 1996, 263, 120.
81. Janin, J. Proteins 1996, 24, R1.
82. Gilson, M. K.; Given, J. A.; Bush, B. L.; McCammon, J. A. Biophys J 1997, 72, 1047.
83. Luo, H.; Sharp, K. A. Proc Natl Acad Sci USA 2002, 99, 10399.
84. Jorgensen, W. L. Acc Chem Res 1989, 22, 184.
85. Jorgensen, W. L.; Buckner, J. K.; Boudon, S.; Tirado–Rives, J. J Chem Phys 1988, 89, 3742.
86. Hermans, J.; Wang, L. J Am Chem Soc 1997, 119, 2707.
87. Roux, B.; Nina, M.; Pomes, R.; Smith, J. C. Biophys J 1996, 71, 670.
88. Leach, A. R. Molecular Modelling—Principles and Applications; Pearson Education Ltd.: Dorchester, 2001.
89. Kollman, P. Chem Rev 1993, 93, 2395.
90. Brady, G. P.; Sharp, K. A. Curr Opin Struct Biol 1997, 7, 215.
91. Murphy, K. P.; Xie, D.; Thompson, K.; Amzel, M.; Freire, E. Proteins 1994, 18, 63.
92. Amzel, L. M. Proteins 1997, 28, 144.
93. Schäfer, H.; Mark, A. E.; van Gunsteren, W. F. J Chem Phys 2000, 113, 7809.
94. Case, D. A. Curr Opin Struct Biol 1994, 4, 285.
95. Levy, R. M.; Karplus, M.; Kushick, J.; Perahia, D. Macromolecules 1984, 1984, 1370.
96. Schäfer, H.; Smith, L. J.; Mark, A. E.; van Gunsteren, W. F. Proteins 2002, 46, 215.
97. Schäfer, H.; Daura, X.; Mark, A. E.; van Gunsteren, W. F. Proteins 2001, 43, 45.
98. Geyer, M.; Schweins, T.; Herrmann, C.; Prisner, T.; Wittinghofer, A.; Kalbitzer, H. R. Biochemistry 1996, 35, 10308.
99. Stumber, M.; Geyer, M.; Graf, R.; Kalbitzer, H. R.; Scheffzek, K.; Haeberlen, U. J Mol Biol 2002, 323, 899.
100. Spoerner, M.; Herrmann, C.; Vetter, I. R.; Kalbitzer, H. R.; Wittinghofer, A. Proc Natl Acad Sci USA 2001, 98, 4944.
101. Vorobjev, Y. N.; Almagro, J. C.; Hermans, J. Proteins 1998, 32, 399.
102. Ladbury, J. E. Chem Biol 1996, 3, 973.