The Use of a Generalized Born Model for the Analysis of Protein Conformational Transitions: A Comparative Study with Explicit Solvent Simulations for Chemotaxis Y Protein (CheY)

MARK S. FORMANECK, QIANG CUI
Theoretical Chemistry Institute, Department of Chemistry, University of Wisconsin, Madison, 1101 University Avenue, Madison, Wisconsin 53706

Received 17 January 2006; Revised 28 March 2006; Accepted 15 April 2006
DOI 10.1002/jcc.20489
Published online 3 October 2006 in Wiley InterScience (www.interscience.wiley.com).

Abstract: To investigate whether implicit solvent models are appropriate for mechanistic studies of conformational transition in proteins, a recently developed generalized Born model (GBSW) was applied to a small signaling protein, chemotaxis protein Y (CheY), with different combinations of the phosphorylation state and conformation of the system; the results were compared to explicit solvent simulations using a stochastic boundary condition. The subtle but distinct conformational transitions involved in CheY activation makes the system ideally suited for comparing implicit and explicit solvent models because these conformational transitions are potentially accessible in both types of simulations. The structural and dynamical properties analyzed include not only those localized to the active site region but also throughout the protein, such as sidechain methyl group order parameters, backbone hydrogen bonding lifetime and occupancy as well as principal components of the trajectories. Overall, many properties were well reproduced by the GBSW simulations when compared with the explicit solvent calculations, although a number of observations consistently point to the suggestion that the current parameterization of the GBSW model tends to overestimate hydrogen-bonding interactions involving both charged groups and (charge-neutral) backbone atoms. This deficiency led to overstabilization of certain secondary structural motifs and more importantly, qualitatively different behaviors for the active site groups (Thr 87, Ala 88, the $\beta4-\alpha4$ loop) in response to phosphorylation, when compared with explicit solvent simulations. The current study highlights the value of carrying out both explicit and implicit solvent simulations for complementary mechanistic insights in the analysis of conformational transition in biomolecules.


Key words: implicit solvent model; generalized Born; conformational transition; chemotaxis; signaling protein

Introduction
Conformational transitions are involved in the function of many biomolecules,1 and hence, revealing the mechanism of such transitions is an important step toward understanding and ultimately manipulating the function and activity of biomolecules. Molecular dynamics simulations have become increasingly powerful as a technique that complements experimental studies in analyzing the pathway, energetics, and regulatory elements of conformational transitions in proteins.2 Because of typical µs–ms time-scale associated with conformational transitions, explicit solvent simulations are usually too computationally demanding to be effective, although notable progress has been made in the cases of small protein and peptide systems.3–7 Accordingly, a substantial amount of research effort has been devoted to developing implicit solvent models8–10 in which the solvent molecules are not explicitly included as discrete particles, but their influence on the solute molecule(s) is described with approximate theories such as the Poisson–Boltzmann approach.11 The use of implicit solvent models provides gains in computational efficiency mainly because of two considerations.

Correspondence to: Q. Cui; e-mail: cui@chem.wisc.edu
Contract/grant sponsor: Research Corporation; contract/grant number: RI1039
Contract/grant sponsor: National Institutes of Health; contract/grant number: RO1-CM071428
Contract/grant sponsor: WARF, University of Wisconsin
This article contains Supplementary Material available at http://www.interscience.wiley.com/jpages/0192-8651/suppmat

© 2006 Wiley Periodicals, Inc.
First, the number of degrees of freedom is substantially reduced compared with that in explicit solvent simulations, which usually have to include a large number of solvent molecules to avoid possible simulation artifacts. Therefore, computing solvent–solute interactions, which is a major computational bottleneck for explicit solvent simulations, can be avoided in implicit solvent simulations. The solvent–solute interaction in implicit solvent models, however, can be fairly complex, and therefore, the practical gain in this aspect can be rather modest (vide infra). Second, because of the neglect of the molecular nature of solvent molecules, implicit solvent simulations can access multiple conformations of the solute at a faster time-scale. Previous studies of peptide folding suggested that implicit solvent simulations speed up conformational transitions by typically a factor of 100 compared with explicit solvent simulations. This is the main reason that implicit solvent models have been used extensively in protein folding and conformational transition problems.

Despite the insights provided by many implicit solvent studies, it has also been recognized that implicit solvent models have limitations. For example, specific protein/DNA–water interactions were observed at the protein–protein or protein–DNA interfaces. Along the same line, salt-bridge interactions were found to be typically overestimated by implicit solvent models, especially in regions where the solvent-separated minimum is expected. These features that reflect the molecular nature of the solvent are difficult to capture with implicit solvent models. Second, almost all the implicit solvent models are at equilibrium in nature; i.e., they assume that solvent molecules reach equilibrium at a much faster time-scale than any major structural reorganization of the solute, which may not always be a valid assumption. For example, the expulsion of water molecules from the protein core was found rate-limiting in the late stage of folding in a number of proteins. Solvent reorganization was also widely recognized to be important in charge separation reactions in the condensed phase such as electron transfer and ion-pair dissociation; significant solvent contribution was also emphasized for peptide isomerization in solution. For these processes, popular implicit solvent models are likely to produce inaccurate results. For example, Chang and Yethiraj found that implicit and explicit solvent simulations gave qualitatively different pathways for the folding of model polymers; implicit solvent simulations tend to get trapped in intermediate states because of the lack of solvent bombardment or too high internal protein–protein friction.

Considering the value as well as fundamental limitations associated with implicit solvent models, it is important to carry out careful comparisons of implicit and explicit solvent simulations, which will establish the problems for which implicit solvent models are appropriate; the results will also provide guidance for the future refinement of these models. Such validation studies were performed by several groups in the past, in which the major issues of concern include whether implicit solvent simulations can maintain the native structure of small proteins and distinguish the native from decoy conformations. More recent benchmark calculations also investigated mean-force, covariance matrix, and main-chain NMR order-parameters for a number of small proteins. An interesting recent study also compared dielectric response of several small proteins from implicit and explicit solvent simulations and found reasonable agreement.

No systematic studies, to the best of our knowledge, have been done to investigate the performance of implicit solvent models in systems with multiple native conformations, such as signaling proteins and biomolecular motors. Considering the potential contribution of implicit solvent simulations to the mechanistic analysis of conformational transition problems in those important biomolecules, we carried out benchmark calculations that compared implicit and explicit solvent simulations using a small signaling protein. Specifically, a recently developed generalized Born model with smooth switching function for the molecular surface (GBSW) was applied to multiple conformational and phosphorylation states of the E. coli chemotaxis Y protein (CheY), which is a representative response-regulator in the two-component signal transduction system in bacteria. More detailed discussions of the activation mechanism of CheY were presented in a separate publication, while the current work focuses on the comparison between results from GBSW and explicit solvent simulations. The subtle but functionally important structural changes in CheY makes it an ideal system for comparing implicit and explicit solvent simulations because these conformational changes are potentially accessible in both types of simulations. The results subject to careful analysis and comparison include not only local structural properties, but also the response throughout the protein to phosphorylation, characterized by quantities such as the main-chain as well as side-chain NMR order parameters. The overall finding is that the GBSW model describes the equilibrium properties of CheY rather well compared with the explicit solvent simulations, although notable differences were also observed that suggest that electrostatic interactions, such as hydrogen bonds involving charged species, are overestimated by the current parameterization of the GBSW model; this trend has also been noted in a few recent studies of several implementations of the GB model. The results also highlighted the value of carrying out both implicit and explicit solvent simulations in the analysis of conformational transition problems, so that complementary mechanistic information can be obtained.

**Computational Methods**

**The Benchmark System: CheY**

The E. Coli chemotaxis Y protein (CheY) is a 129-residue (14 kDa) signal transduction protein that is activated through phosphorylation. Upon phosphorylation at Asp 57, a tyrosine residue (Tyr 106) ~10 Å away switches from being solvent-exposed to a buried cavity under the β4–α4 loop (Fig. 1). It is generally believed that the isomerization of Tyr 106 modulates the binding affinity of CheY to FlIm, which in turn influences the rotational pattern of the flagella (clock-wise vs. counter-clock-wise rotation) and, consequently, the chemotaxis behavior of bacteria (tumbled vs. smooth swimming). The two crystal structures used in this study represent the inactive (PDB:1JBE, Chain B, 1.1 Å resolution) and active (PDB:1F4V, Chain A, 2.2 Å resolution) states of CheY; the active structure was obtained using beryllium trifluoride as the phosphate mimic, which exhibits similar activity as that of phosphate, yet does not dissociate as quickly. Both crystal structures omit the first methionine residue, and therefore, the simulated system
Comparison of Generalized Born Model with Explicit Solvent Simulations for CheY

Figure 1. The active-site of CheY in (a) the unphosphorylated-inactive and (b) the phosphorylated(with Mg$^{2+}$)-active states based on the X-ray structures.\textsuperscript{48, 51} The side-chain of Tyr 106 undergoes an isomerization as a functional response to the activation. It was proposed\textsuperscript{41} that phosphorylation of Asp 57 stabilizes two hydrogen-bonding interactions involving Thr 87 and Ala 88 and therefore the active configuration of the $\beta_4-\alpha_4$ loop, which in turn modulates the rotameric state of Tyr 106. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

was 128 residues in length. To correlate the phosphorylation event and the conformational properties of CheY, both structures were simulated with and without the phosphate group and the accompanying magnesium ion at the active site (Asp 57). The expectation was for the phosphorylated-inactive and unphosphorylated-active structures to show signature of structural instability, as there may be tendencies for conformational changes to occur because the phosphorylation state is perturbed.

As discussed in detail in previous studies,\textsuperscript{41, 46, 47} in addition to the isomerization of Tyr 106, CheY displays subtle conformational changes upon phosphorylation (Fig. 1). The most relevant structural properties involve the hydrogen-bonding distances from Asp 57 to the hydroxyl group of Thr 87 and to the backbone amine of Ala 88, as well as the conformation of the nearby $\beta_4-\alpha_4$ loop (residues Ala 88 to Lys 91).\textsuperscript{51} Our recent simulation study\textsuperscript{41} supported an activation mechanism in which the phosphorylation of Asp 57 modulates the dominant configuration of the $\beta_4-\alpha_4$ loop through the two key hydrogen-bonding interactions mentioned earlier, and the loop configuration in turn regulates the rotameric state of Tyr 106. Since isomerization of Tyr 106 is too rare to observe in unbiased molecular dynamics simulations,\textsuperscript{41} here we focus on the hydrogen-bonding pattern of Asp 57 and the $\beta_4-\alpha_4$ loop configuration in implicit and explicit solvent simulations. As emphasized in Introduction, the local nature of these functionally important structural properties makes CheY an ideal system for contrasting implicit and explicit solvent simulations, because responses of those properties to phosphorylation are likely accessible in both types of simulations.

 Explicit Solvent Simulation: The Stochastic Boundary Condition

For the explicit solvent simulations, the stochastic boundary condition\textsuperscript{54, 55} (SBC) was applied, in which the protein was solvated by a water droplet of 30 Å radius. Water molecules were subject to a deformable boundary potential\textsuperscript{54} starting at 26 Å. To simulate the random bombardment and viscosity of the missing solvent past 30 Å, Langevin dynamics\textsuperscript{56} was performed for water molecules beyond 27 Å from the center of the droplet, for which a friction coefficient of 80 ps$^{-1}$ (on the oxygen atoms only) and a heat bath of 300 K were used. Protein atoms and other solvent molecules were treated with the regular Newtonian dynamics.

The protein atoms were described with the CHARMM 22 all-atom force field,\textsuperscript{57} while the water molecules were described with a modified version\textsuperscript{58} of the TIP3P model.\textsuperscript{59} Parameters for the phosphate group were taken from the CHARMM 27 lipid parameter set,\textsuperscript{60} and those for the magnesium ion were taken from the CHARMM 27 force field.\textsuperscript{61} van der Waals interactions were calculated for atoms within 8 Å and switched to zero at 12 Å with a smooth switching function.\textsuperscript{56, 62} Electrostatic interactions were calculated based on the extended electrostatics scheme,\textsuperscript{56, 63} in which interactions among groups beyond 13 Å were calculated with a multipolar expansion truncated at the quadrupole level.

To solvate the protein, a pre-equilibrated 30 Å radius water droplet was overlaid on the protein, and water molecules with the oxygen atom within 2.8 Å of any nonhydrogen protein atom were deleted. The protein was held fixed, while the positions of the water
molecules were minimized using 2000 steps of the steepest descent method\textsuperscript{64} and then equilibrated with 5 ps of Nose-Hoover molecular dynamics\textsuperscript{65,66} at 300 K. Resulting cavities were filled by overlaying the system with three additional water spheres, each rotated 90° in the x, y, and z direction, respectively. Water molecules overlapping with the protein or previously placed water were deleted. All water molecules were further minimized and then relaxed with 10 ps of Nose-Hoover molecular dynamics. Afterward, the entire system was relaxed without any constraints with 500 steps of steepest descent minimization followed by 1500 steps of Adopted Basis Newton-Raphson minimization.\textsuperscript{56} The final system included about 3250 water molecules and 1981 (1986) protein atoms for the unphosphorylated (phosphorylated) CheY.

During the equilibration and production molecular dynamics simulations, the protein was restrained to the center of the water droplet, using a weak harmonic potential applied to the protein center of mass (calculated based on the backbone atoms only) with a force constant of 10.0 kcal/(mol Å\textsuperscript{2}). The leap-frog integration scheme\textsuperscript{67} was used with the 2-fs time-step, and the SHAKE algorithm\textsuperscript{68} was applied to constrain all bonds involving hydrogen. For heating, the system temperature was raised from 50 to 300 K with 50 ps of molecular dynamics during which velocities were rescaled every 2 ps to increase the temperature by 10 K. The overall translation and rotation of the system was stopped every 1 ps. The temperature was checked every 1 ps and was allowed to fluctuate within 10 K of the target temperature; otherwise, the velocities were rescaled. Following heating, the system was equilibrated at 300 K for an additional 50 ps. For production runs, approximately 7 ns simulations were carried out for each trajectory. For each combination of conformational (inactive vs. active) and phosphorylation state, two separate trajectories (using different initial random seeds for velocity assignment) were carried out, which produced a total of eight explicit solvent trajectories (see Table 1).

**Implicit Solvent Simulation: The GBSW Model**

For the implicit solvent simulations, the generalized Born with smooth switching function (GBSW) method\textsuperscript{66} was used. This model was chosen because it was recently developed and was shown to provide a reasonable balance between computational efficiency and accuracy.\textsuperscript{37} In the following paragraphs we briefly review the key elements of this model to facilitate subsequent discussions regarding parameters used in the GBSW simulations and the comparison with explicit solvent simulations.

Like most implicit solvation models,\textsuperscript{69} the GBSW model partitions the solvation free energy of a macromolecule into the electrostatic and nonpolar contributions.

\[
\Delta G_{\text{solv}} = \Delta G_{\text{solv,elec}} + \Delta G_{\text{solv,nonp}}.
\]

Concerning the electrostatic component, all generalized Born (GB) solvation models theoretically start from the premise that the work required to charge up a point charge, \(q_i\), in the center of a sphere of radius \(R_i\) embedded in a dielectric medium of dielectric constant \(\varepsilon_r\), is given by the Born model:\textsuperscript{70}

\[
\Delta G_{\text{solv,elec}} = -\frac{1}{2} \left( \frac{1}{\varepsilon_p} - \frac{1}{\varepsilon_r} \right) \frac{q_i^2}{R_i}
\]

where \(\varepsilon_p\) is the dielectric constant inside the sphere.

GB methods extend this result to approximate the electrostatic solvation free energy for a set of point charges placed in an arbitrarily shaped cavity. The corresponding equation for the electrostatic solvation free energy of a macromolecule given by Still et al. is:\textsuperscript{71}

\[
\Delta G_{\text{solv,elec}} = -\frac{1}{2} \left( \frac{1}{\varepsilon_p} - \frac{1}{\varepsilon_r} \right) \sum_{\eta} \frac{q_i q_j}{\sqrt{r_{ij}^2 + \alpha_i \alpha_j \exp(-r_{ij}^2/4\alpha_i \alpha_j)}}
\]

where the “effective” or “generalized” Born radii, \(\alpha_i\), depend on the position of the atom that bears the charge \(q_i\).

Assuming that the potential in the dielectric medium follows Coulomb’s law, an estimate for the generalized Born radii is given by the Coulomb field approximation (CFA),\textsuperscript{72}

\[
\frac{1}{\alpha_i} = A_4 = \frac{1}{R_i} = \frac{1}{4\pi} \int_{\text{solute,r} > R_i} \frac{1}{r^2} \, dV
\]

where \(R_i\) is the position-independent atomic radius. The CFA neglects the reaction field because of a true dielectric medium; as a result, the estimated Born radii were too large, while the estimated solvation free energies tend to be too small.\textsuperscript{73} In the GBSW model, improvements are made by including a higher order correction term:\textsuperscript{73}

\[
A_7 = \left( \frac{1}{4\pi R_i^2} - \frac{1}{4\pi} \int_{\text{solute,r} > R_i} \frac{1}{r^2} \, dV \right)^{1/4}
\]

which is combined with \(A_4\) in eq. (4) to give an estimate for the Born radii in the following manner:

\[
\alpha_i = \frac{1}{C_0 A_4 + C_1 A_7}
\]
where the parameters $C_0$ and $C_1$ are parameterized for a particular switching width during which the dielectric constant is switched smoothly from $\varepsilon_p$ to $\varepsilon_s$ across the protein surface. The numerical quadrature techniques used for integration in eqs. (4) and (5) prove to make GBSW more efficient than previous implementations of GB models, making GBSW more useful for molecular dynamics simulations.

In the current work, the set of atomic radii derived by Nina and coworkers was used for $R_i$ in eqs. (4) and (5), while the standard van der Waals radii were used for phosphate and Mg$^{2+}$; for hydrogen, the convention of $R_i = 0$ was adopted. For the dielectric switching, a smoothing length of 0.4 Å was used. For the characterization of the protein surface, the sharp “molecular surface” was found to produce more stable MD trajectories (in terms of overall RMSD) than the smooth protein surface based on $\sim$1 ns MD simulations (see Supplementary Materials) and therefore was chosen for production simulations. For the molecular surface, the corresponding Born radii coefficients are $C_0 = 1.204$ and $C_1 = 0.187$. The effective Born radii were updated every other molecular dynamics or minimization step for computational efficiency.

As to the nonpolar contribution to solvation free energy, the standard model based on the solvent accessible surface-area (SASA) was used,

$$\Delta G_{\text{solv,sp}} = \gamma \times \text{SASA} \quad (7)$$

where $\gamma = 0.03 \text{ kcal/(mol } \text{Å}^2\text{)}$ was used for this study. This value was used in the original work that developed GBSW and was further tested by comparing the behavior of the $\beta_4-\alpha_4$ loop in short (1.5 ns) GBSW simulations and explicit solvent simulations: this loop was chosen because its different configurations in different conformational states of CheY have substantially different SASA. Using the value of 0.03 kcal/(mol Å$^2$) for $\gamma$ was found to give very similar $\beta_4-\alpha_4$ loop configuration compared to explicit solvent simulations for the unphosphorylated-inactive state. When compared with the value of 0.005 kcal/(mol Å$^2$) used in previous studies, the current choice is likely a bit high and contributes to the observed over-compactation found in the GBSW simulations (see later); most other properties, however, were observed to be rather insensitive to the choice of $\gamma$ in the range of 0.01 and 0.03 kcal/(mol Å$^2$) (see Supplementary Material).

The magnesium ion, which was present in phosphorylated simulations, was expected to coordinate to both the protein and water molecules. Since no explicit water was included in the GBSW simulations, constraints were found necessary to maintain the magnesium ion stable in the active site. Specifically, NOE constraints were added for distances between Mg$^{2+}$ and four neighboring atoms: phosphorus and the O$_\text{y1}$ of Asp 57, the backbone carbonyl oxygen of Arg 59, and C$_\gamma$ of Asp 13.

To be compatible with the way that the GBSW model was parameterized, large nonbonded cutoff schemes were considered. Test calculations compared with simulations without any cutoff suggested that a switching scheme with a cutoff of 20 Å and energy-switching between 18 and 20 Å produced the best balance of stable simulation and computational efficiency. Because of the rather large cutoff, the GBSW simulations were only faster than the explicit solvent simulations by a factor of two. However, as emphasized in Introduction, the major gain in implicit solvent simulations is expected to originate from the increase in the efficiency of conformational sampling.

To account for the lack of explicit collision with the solvent molecules, Langevin dynamics was carried out for all GBSW simulations at 300 K with a simple scheme in which all nonhydrogen protein atoms were assigned with a friction coefficient of 25 ps$^{-1}$; although a value of 50 ps$^{-1}$ has been found to substantially decrease the mobility of proteins, the properties of interest here were found to be insensitive to the value of the friction coefficient in the range of 10–25 ps$^{-1}$ (data not shown). Similar to the explicit solvent simulations, the CHARMM 22 force field was used to describe the protein atoms, and the SHAKE algorithm was used to constrain all bonds involving hydrogen to allow an integration time step of 2 fs. As shown in Table 1, at least two GBSW trajectories were calculated for each combination of conformational and phosphorylation states. Each trajectory was typically run for 9 ns, and much longer (up to 100 ns) simulations were carried out for several trajectories.

### Analysis Procedures

Standard analyses were performed on the calculated trajectories, which include root-mean-squared deviation (RMSD) relative to the X-ray structures, root-mean-squared fluctuations (RMSF) and the radius of gyration ($R_g$). In addition, a few other more sophisticated analyses were carried out to better compare the implicit and explicit solvent simulations. These include main-chain and side-chain NMR order parameters, backbone hydrogen-bonding lifetime/occupancy and the quasi-harmonic analysis; technical details for those analyses are summarized here.

#### $S^2$ and $S_{\text{axis}}^2$ NMR Order Parameters

The NMR order parameters provide a measure of the average spatial extent for the motion of a bond vector and therefore characterize the internal flexibilities of local structural motifs in a biomolecule. They complement other measures such as the RMSF because measurements on side-chain order parameters are governed by both main-chain fluctuations as well as side-chain dihedral rotations. Similarly, measurements on the backbone may be slightly complicated by backbone distortions. Because of the lack of experimental NMR order parameter data for CheY, we restrict ourselves to comparing results from implicit and explicit solvent simulations, which fits the purpose of the current work. Both the main-chain N–H order parameters ($S^2$) and the side-chain methyl axial rotation order parameters ($S_{\text{axis}}^2$) were calculated because increasing amount of evidence has indicated that main-chain and side-chain motions may change rather differently during a structural transition.

Within the Szabo-Lipari framework, the order parameter is given by the long time limit of the time-correlation function that involves the second-order Legendre polynomial, $P_2$, of the bond-vector orientations ($\tilde{\mu}$) at different times,

$$S^2 = \lim_{t \to \infty} \langle P_2(\tilde{\mu}(t) \cdot \tilde{\mu}(\tau + t)) \rangle \quad (8)$$
For the methyl axial rotation order parameter, experimentally determined side-chain order parameters are usually derived from measurements of the C–H bonds in the methyl group, where $S_{\text{ax}}^2 = S^2/0.111$. In the current study, however, the C–C bond was explicitly followed, because of the observation that the methyl rotation may not be free enough to produce converged results due to an overestimate of side-chain rotational barriers in the current CHARMM force field.\cite{84, 85}

From the computational perspective, an alternative “equilibrium expression” for the order parameter is often found convenient:\cite{86}

$$S^2 = \frac{3}{2} \left( \langle x^2 \rangle^2 + \langle y^2 \rangle^2 + \langle z^2 \rangle^2 + 2\langle xy \rangle^2 + 2\langle xz \rangle^2 + 2\langle yz \rangle^2 \right) - \frac{1}{2}$$

(9)

where $x$, $y$, and $z$ indicate different components of the bond orientation vector $\vec{\mu}$. For fully converged samplings, both the time-correlation (eq. 8) and the equilibrium (eq. 9) expressions give the same result (see discussion in Refs. 87 and 88). However, since less data are used in the time-correlation approach, the equilibrium approach was preferred in many previous studies\cite{86} as well as here. Specifically for the current system, we found that both approaches gave similar backbone $S^2$ values, but better converged values were obtained via the equilibrium approach for the side-chain order parameters.

**Backbone Hydrogen Bonding Analysis**

The properties of hydrogen bonds involving backbone atoms were used as an useful observable in a previous benchmark study of GB
models. It was found that the number of backbone hydrogen bonds maintained during the simulation differed significantly between different implicit solvent models. Motivated by that work, we compared the average lifetime and occupancy of backbone hydrogen bonds in explicit solvent and GBSW simulations.

A backbone hydrogen bond was counted if the distance between the amine hydrogen and carbonyl oxygen was less than 2.4 Å and the N–H – O angle was between 140° and 180°. The occupancy of hydrogen bonding was assigned to the residue containing the hydrogen based on the fraction of snapshots in which a hydrogen bond was formed versus the total number of snapshots analyzed: snapshots were analyzed every 1 ps during the production run. The lifetime of a hydrogen bond was defined as the average time between its formation and rupture during the molecular dynamics simulation. The results were averaged over all independent runs for each simulation condition.

**Quasiharmonic Normal Modes**

In addition to those local properties discussed earlier, it is interesting to investigate whether implicit solvent models reproduce collective, global properties of proteins. This could be particularly important in allosteric systems in which long-range correlation is important for function. To this end, coordinate covariance matrix, \( C_{ij} \), was analyzed.

\[
C_{ij} = \langle q_i q_j \rangle
\]

where the mass-weighted cartesian displacements \( (q_i) \) were used.

The covariance matrix was then diagonalized, which produced quasiharmonic modes or principal components, \( U_i \), that provide a quantitative description for the collective motion of the protein. To compare the principal components from different simulations,
two types of measures were made. First, the overlap matrix, $M$, between quasiharmonic modes from different simulations was computed:

$$M_{ij} = U_i^T U_j.$$  \hspace{1cm} (11)

In addition, it is of interest to measure how the entire subspace involving collective motions is represented by different simulations. For this purpose, the “spanning coefficients” were calculated between pairs of simulations,

$$\text{SPAN}_i = \sum_j (U_i^T U_j)^2 = \sum_j (M_{ij})^2$$  \hspace{1cm} (12)

where the summation in $j$ was limited to the first 250 modes (CheY has $\sim 6000$ internal modes), which roughly covers 1–40 cm$^{-1}$ range.

Considering the cost of diagonalizing a rather large covariance matrix, the quasiharmonic analysis comparison was only carried out for one state; the unphosphorylated-inactive state was chosen because the GBSW and the explicit solvent simulations gave, overall, the most similar structural properties. The comparison was made between different trajectories with the same solvent model but different initial conditions as well as between trajectories, using different solvent models.

### Results and Discussion

Using CheY as an example, we compare implicit and explicit solvent simulations to better establish whether implicit solvent models,
Comparison of Generalized Born Model with Explicit Solvent Simulations for CheY

Figure 5. A snapshot of the $\beta_4-\alpha_4$ loop region in the GBSW 1 simulation for the phosphorylated-inactive state after 50 ns, when compared with the crystal structures of the inactive (blue) and active (yellow) states of CheY. An aberrant hydrogen bond between the Ala 90 amine and the Thr 87 hydroxyl group led to a rather distorted configuration for the $\beta_4-\alpha_4$ loop, which had a large ($\sim$5 Å) RMSD relative to both the inactive and active X-ray data; this unexpected loop configuration lasted for more than 60 ns in the GBSW simulation.

such as the GBSW model used here, are likely appropriate for studying conformational transitions in biomolecules. Since such transitions, especially in allosteric systems, typically involve a localized activation event (e.g., ligand binding or phosphorylation) as well as more collective structural changes, properties that are both localized to the active site(s) and throughout the protein in response to the activation event need to be compared. In the following discussion, we first focus on the properties of the phosphorylation and response sites and then move on to other properties throughout the protein, which include global structure and fluctuations, NMR order parameters, backbone hydrogen-bonding patterns, and principal components.

Active-Site Properties

In CheY, phosphorylation of Asp 57 induces subtle but crucial conformational changes at the binding site for the flagella protein FliM. The most notable change concerns the rotation of the Tyr 106 side-chain more than 10 Å away from Asp 57.48 Two key hydrogen-bonding interactions involving the phosphorylated Asp 57 were proposed by both experiments51 and our recent simulations41 to be important for the activation; these interactions (Fig. 1) involve the side-chain of the conserved Thr 87 and the main-chain NH of Ala 88. Briefly, these hydrogen-bonding interactions, which are formed and remain stable when Asp 57 is phosphorylated, are proposed41 to regulate the configuration of the $\beta_4-\alpha_4$ loop, which in turn influences the rotameric state of Tyr 106. Accordingly, local properties of crucial functional importance, which are discussed later, include the hydrogen-bonding patterns involving Thr 87 and Ala 88 as well as the flexibility of the $\beta_4-\alpha_4$ loop.

Hydrogen Bonds Involving Key Active-Site Residues

Since the hydrogen bonds involving Thr 87 and Ala 88 play an important regulatory role in CheY activation, the formation and breaking of these interactions with Asp 57 are expected to depend on the phosphorylation state of the protein. The specific behavior of these crucial hydrogen-bonding interactions in the simulations, however, was found to depend on whether the solvent was treated explicitly or implicitly.

Among the four states considered, the UnP-Inact and P-Act states were expected to be stable, since the phosphorylation state was not perturbed from that of the corresponding crystal structures. This was indeed observed in both explicit solvent and GBSW simulations, where the Thr 87–Asp 57 and Ala 88–Asp 57 hydrogen bonds were not formed in the UnP-Inact simulations (Figs. 2a and 3a), but remained fairly stable in the P-Act simulations (Figs. 2d and 3d). The only moderate exception involved the hydrogen bond between Ala 88 and phosphorylated Asp 57 in the P-Act state, which was found to constantly break and form in both explicit solvent trajectories, but rarely went beyond 3 Å in the GBSW simulations (Fig. 3d).

When the phosphorylation state was perturbed, by either adding a phosphate to the inactive conformation or removing the phosphate from the active conformation, more significant differences were found between explicit solvent and GBSW simulations. For the P-Inact state, both Thr 87 and Ala 88 formed hydrogen-bonds...
Figure 6. Root mean squared fluctuations for Cα atoms averaged over all simulations for each state and solvent model; see Table 1 for notation. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

with the phosphorylated Asp 57 in multiple GBSW simulations, while this was not observed in any explicit solvent simulations (Figs. 2b and 3b). For the UnP-Act state, the hydrogen bonds between Thr 87/Ala 88 and Asp 57 remained largely stable in the GBSW simulations but both eventually broke in all explicit solvent simulations (Figs. 2c and 3c).

These observations can be readily explained by the conjecture that the current parameterization of the GBSW model overestimates the strength of hydrogen-bonding interactions. Since the overall protein structure is well maintained (see the next section), it is likely that such overestimation is severe particularly for interactions involving charged groups (in this case the side-chain of Asp).

Table 2. The RMSD of the Average Structure Relative to the X-ray Structure and the Radius of Gyration ($R_g$) from Different Simulations of CheY in Different Chemical and Conformational States.a

<table>
<thead>
<tr>
<th>State</th>
<th>SBC-RMSD (Å)</th>
<th>GBSW-RMSD (Å)</th>
<th>SBC-$R_g$ (Å)</th>
<th>GBSW-$R_g$ (Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td>UnP-Inact</td>
<td>1.2 ± 0.1</td>
<td>1.1 ± 0.1</td>
<td>13.62 ± 0.05</td>
<td>13.46 ± 0.04</td>
</tr>
<tr>
<td>P-Inact</td>
<td>1.3 ± 0.2</td>
<td>1.2 ± 0.1</td>
<td>13.66 ± 0.06</td>
<td>13.46 ± 0.04</td>
</tr>
<tr>
<td>UnP-Act</td>
<td>1.6 ± 0.2</td>
<td>1.4 ± 0.1</td>
<td>13.76 ± 0.08</td>
<td>13.58 ± 0.06</td>
</tr>
<tr>
<td>P-Act</td>
<td>1.8 ± 0.3</td>
<td>1.9 ± 0.1</td>
<td>13.68 ± 0.07</td>
<td>13.56 ± 0.06</td>
</tr>
</tbody>
</table>

*aSee Table 1 for the notation used to characterize different chemical-conformational states. SBC and GBSW indicate explicit solvent and implicit solvent simulations carried out here, respectively. The RMSD and $R_g$ values were calculated for the last 6 and 8 ns of the explicit solvent and GBSW simulations, respectively (i.e., after the first nanosecond of production simulation).
Comparison of Generalized Born Model with Explicit Solvent Simulations for CheY

Figure 7. The side-chain methyl rotation order parameters ($S^2_{axis}$) averaged over all simulations for each state and solvent model (see Table 1 for a summary); for the P-Inact state, only 18 ns data were included from the 100 ns GBSW simulation because the $\beta_4$–$\alpha_4$ loop configuration became too distorted in the later part of the simulation (see Fig. 5 and discussion). Secondary structures are indicated in each graph. See Table 1 for notation. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

57 and phosphate; however, also see the discussions in Backbone Hydrogen Bonding). This overestimation is likely due to the lack of explicit water molecules in the first solvation shell of (particularly) charged groups, which effectively weaken hydrogen-bonding or salt-bridge interactions by providing alternative coordinations. For example, the hydrogen bond between Thr 87 and unphosphorylated Asp 57 was found to constantly break in explicit solvent simulations because water molecules were able to coordinate with either the OH in Thr 87 or the carboxylate in Asp 57. This rather general limitation of implicit solvent models has been discussed by several authors in the past.\textsuperscript{22, 97} e.g., the occurrence of direct hydrogen bonds between phosphate and peptide backbone in small peptide simulations was found significantly increased with an implicit solvent model (another generalized Born model) when compared with explicit solvent simulations.\textsuperscript{97} Yu et al. proposed to include one explicit water molecule to better describe salt-bridges in implicit solvent simulations.\textsuperscript{22} Alternatively, it is possible to re-scale the atomic radii to achieve a balanced treatment of electrostatic interactions between charged or polar groups in the presence of solvation.\textsuperscript{30, 42, 43}

$\beta_4$–$\alpha_4$ Loop Configuration

The conformation and flexibility of loops are of particular interest because of the diverse roles of loops in protein functions.\textsuperscript{98–100} In CheY, the $\beta_4$–$\alpha_4$ loop, which consists of residues Ala 88–Lys 91, was observed to adopt different configurations in the inactive\textsuperscript{51} and active\textsuperscript{48} X-ray structures. It was proposed\textsuperscript{41, 51} that the configuration of this loop might gate the isomerization of Tyr 106. To characterize the configuration of this central loop in the simulation, we monitored the differential RMSD with respect to the inactive and active loop configurations in the corresponding X-ray structures: $\Delta$RMSD($t$) = RMSD($X(t)$; active) – RMSD($X(t)$; inactive); a positive $\Delta$RMSD value indicates a $\beta_4$–$\alpha_4$ loop configuration closer to the inactive X-ray data.

The $\beta_4$–$\alpha_4$ loop configuration is largely regulated by the hydrogen bonding interactions between Thr 87/Ala 88 and Asp 57; when
Table 3. The Axial Methyl Rotation Order Parameters \( (S_{2ax}^m) \) for a Few Side-Chain Methyl Groups in the Phosphorylated-Inactive State with Different Solvation Models and Trajectories.

<table>
<thead>
<tr>
<th>Run</th>
<th>Time of analysis (ns)</th>
<th>Ile 55</th>
<th>Met 85</th>
<th>Thr 87</th>
</tr>
</thead>
<tbody>
<tr>
<td>SBC 1</td>
<td>7</td>
<td>0.84</td>
<td>0.17</td>
<td>0.86</td>
</tr>
<tr>
<td>SBC 2</td>
<td>7</td>
<td>0.60</td>
<td>0.28</td>
<td>0.88</td>
</tr>
<tr>
<td>GBSW 1</td>
<td>100</td>
<td>0.19</td>
<td>0.21</td>
<td>0.86</td>
</tr>
<tr>
<td>GBSW 2</td>
<td>9</td>
<td>0.43</td>
<td>0.82</td>
<td>0.90</td>
</tr>
<tr>
<td>GBSW 3</td>
<td>9</td>
<td>0.28</td>
<td>0.53</td>
<td>0.86</td>
</tr>
</tbody>
</table>

The axial methyl rotation order parameters were calculated based on (eq. (9)). For Ile 55, the values shown are for the C\( \gamma \)–C\( \delta \) bond vector. Ile 55 and Met 85 are in the core of CheY (see Fig. 8) and Thr 87 is between the phosphorylation site (Asp 57) and response site (Tyr 106).

both hydrogen bonds are formed, this loop is held close to the active configuration, while as any one of the interactions is unstable or broken, the fluctuation in the loop increases. Therefore, the way that the loop responded to phosphorylation in different simulations followed a similar pattern as observed for the key hydrogen bonds in the phosphorylation site. The UnP-Inact state was the only one for which explicit solvent and GBSW simulations gave very similar results; i.e., the loop remained essentially inactive, as reflected by the largely positive \( \Delta R_{M\text{S}} \) values (Fig. 4a), in all simulations. For all other states, the behavior of the loop was quite different in GBSW and explicit solvent simulations. For P-Act, the hydrogen bond between the phosphorylated Asp 57 and Ala 88 was found to constantly break and form in explicit simulations, but remained steady in GBSW simulations (Figs. 2d and 3d); the difference in this single hydrogen bond was sufficient to cause the \( \beta_4-\alpha_4 \) loop to oscillate between the inactive and active configurations in the explicit solvent simulations but remain very close to the active configuration in the GBSW simulations (Fig. 4d). Also parallel to the behavior of key hydrogen bonds, the \( \beta_4-\alpha_4 \) loop underwent a major structural transition in the P-Inact state only with the GBSW simulations (Fig. 4b), while in the UnP-Act state only with explicit solvent simulations (Fig. 4c).

An unexpected loop configuration was observed in one of the GBSW simulations for the P-Inact state, which was carried out

Figure 8. Two residues in the protein core, Ile 55 and Met 85, which were found to have consistently low \( S_{2ax}^m \) values in all simulations (see Table 3), are shown in the inactive CheY crystal structure (PDB 1JBE).\(^{50}\) [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]
Comparison of Generalized Born Model with Explicit Solvent Simulations for CheY

Figure 9. The convergence behavior of \( S_2 \) axis for the side-chain of Thr 87 and relevant structural properties in the phosphorylated-inactive state with GBSW simulation no. 1. (a) Cumulative average of the equilibrium expression of \( S_2 \) axis (eq. (9)); (b) the \( \chi_1 \) side-chain dihedral angle; (c) distance from Thr 87 O\(^\gamma_1\) to the closest Asp 57 phosphate oxygen. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

for 100 ns with the hope that unbiased isomerization of Tyr 106 can be observed. However, as illustrated by the snapshot shown in Fig. 5, the loop became substantially distorted from the configurations found in either the inactive or active X-ray structures: a hydrogen bond was formed between the hydroxyl group of Thr 87 and the carbonyl of Ala 90, and the RMSD of the loop was as large as 5 Å from both the inactive and active conformations. This hydrogen-bonding pattern was not observed in any solved X-ray structure of CheY and did not break within 60 ns following its formation, which suggest that the result is likely an artifact of the GBSW model. Apparently, the enhancement of hydrogen bonding interactions may lead to the sampling and over stabilization of rare conformations.

Global Protein Properties

In many proteins, especially those involved in signal transduction and energy conversion, local events induce significant changes throughout the system. Therefore, in addition to examining functionally important local interactions, it is of interest to analyze whether the properties of the protein are globally well-described by implicit solvent models.

RMSF, RMSD, and the Radius of Gyration

Overall, the RMSF from explicit solvent and GBSW simulations agree well for all the four states simulated (Fig. 6). For example, regions of high RMSF correspond to various loop regions and the two termini, although the GBSW simulations were observed almost consistently to produce slightly lower RMSF values. Some of the differences are related to the hydrogen-bonding pattern as discussed earlier. For the \( \beta_4-\alpha_4 \) loop in the P-Act state, for example, the loop RMSF in the explicit solvent simulations were substantially higher (Fig. 6d) because the hydrogen bond between Ala 88 and the phosphorylated Asp 57 could be replaced by water molecules, whereas this was not possible in the GBSW simulations. For the UnP-Act state, the loop around residue 70 (see Fig. 8) showed much higher fluctuations in one of the explicit solvent simulations (Fig. 6c) because the intraprotein hydrogen bonds that maintained the stability of the loop were replaced by interactions with the solvent after about 6 ns into the simulation; because of the lack of explicit solvent, the GBSW simulations maintained all intraprotein hydrogen bonds in this region throughout the two trajectories that lasted 24 and 15 ns, respectively, and hence produced substantially lower fluctuations.
The overall RMSD’s of the average structure from explicit solvent and GBSW simulations relative to the relevant X-ray structures were rather close for all four states studied (Table 2). However, it has been shown that the overall RMSD may not be a sensitive measure for the quality of the solvation model compared with other observables, such as the radius of gyration ($R_g$) and number of native hydrogen bonds maintained in the simulation. As shown in Table 2, the difference in $R_g$ between explicit solvent and GBSW simulations was more substantial when compared with the corresponding fluctuation. The GBSW model tends to give lower $R_g$ values, which imply more compact structures; this is also consistent with the above conjecture that the current parameterization of GBSW tends to overestimate the strength of hydrogen bonding interactions, although the relatively high value of surface tension adopted here also likely contributes to the decreased $R_g$ (see Supplementary Material).

**NMR Order Parameters**

The backbone NH order parameters (see Supplementary Materials) showed no apparent solvation model dependence in all CheY states studied here. The overall trends were highly antiparallel to those for the C$_\alpha$ RMSF values shown earlier, i.e., lower $S^2$ values correspond to higher RMSF values. The side-chain methyl order parameters, $S_{\text{axis}}^2$, which have been documented to provide complementary information regarding protein dynamics during conformational transitions (also see Supplementary Materials), showed a richer degree of variations with different states and solvent models.

One critical issue in the calculation of side-chain order parameters is convergence. As discussed in the literature, the convergence of $S_{\text{axis}}^2$ may take a very long simulation time due to the rare nature of side-chain isomerization among different rotameric states (also see later). In this regard, it has been suggested that implicit solvent simulations, which generally allow a higher degree of mobility for side-chains, may lead to lower $S_{\text{axis}}^2$ parameters compared to explicit solvation models. In the current study, as shown in Figure 7, little consistent difference was observed between the explicit solvent and GBSW simulations (see Supplementary Materials), except in the UnP-Inact state for which the explicit solvent runs produced consistently lower $S_{\text{axis}}^2$ values. This was curiously opposite to previous observations based on computer simulations. Further analysis (not shown) revealed that this finding was the same for both buried and exposed residues.
One interesting similarity found for all states and solvation models was for Met 85 and Ile 55 (Cγ–Cδ), which had mostly low $S^2_{\text{axis}}$ values (see Table 3) despite their presence in the protein core (Fig. 8) and the low flexibility of the surrounding backbone atoms (according to RMSF and NH $S^2$). This observation demonstrated the flexibility of side-chains in the protein core, as also suggested in previous studies. The fact that similar $S^2_{\text{axis}}$ values were found in the explicit solvent and GBSW simulations for these core residues is significant in the current context considering the above discussion that GBSW tends to produce too compact structures. In other words, it seems that the dynamical consequence of such compaction was not significant and the mobility of protein groups were, in fact, largely well-reproduced by the GBSW model.

Another interesting case concerns Thr 87, whose side-chain order parameter converged rather well in all simulations, since the backbone and the side-chain did not fluctuate significantly. As shown in Figure 9 with the very long GBSW simulation, however, sharp change in the cumulative average (eq. (9)) occurred when the side-chain $\chi_1$ angle was stable (Fig. 9b), but backbone atoms underwent displacements (Fig. 9c). The same effect was observed for alanine residues, which had no axial flexibility, but exhibited poor convergence in $S^2_{\text{axis}}$ due to flexibility in the backbone. Hence $S^2_{\text{axis}}$ order parameters can be helpful for monitoring conformational heterogeneity (when converged), yet the origin of the flexibility may be hard to elucidate. Comparison with backbone NH $S^2$ parameters may help to discern between backbone and side-chain flexibilities. Despite such subtleties, as seen in Table 3, the Thr 87 side-chain order parameters were consistent between solvent models and between multiple simulations.

Finally, it is useful to illustrate the convergence issue for $S^2_{\text{axis}}$ calculations with different solvent models. For this purpose, the convergence behavior of $S^2_{\text{axis}}$ and relevant structural parameters from multiple simulations are plotted for Val 108 in Figures 10–12. The methyl group in the explicit solvent simulations exhibited very little rotation about the $\chi_1$ dihedral. As a result, the $S^2_{\text{axis}}$ values in these simulations were very high (Fig. 10). With the GBSW model, the Val 108 methyl exhibited much more frequent rotations between $-60^\circ$ (when both methyl groups were oriented toward Ala 90) and $180^\circ$ (when one of the methyl groups was oriented towards Ala 90/Tyr 106).
and the other pointed away) in one of the simulations (GBSW 3, Fig. 11). Whenever a rotamer exchange took place, the cumulative average of the order parameter was significantly affected. Correspondingly, $S^2_{\text{axis}}$ converged to a low value within 9 ns in the GBSW run 3 (Fig. 11), while it took about 50 ns for the GBSW run 1 (Fig. 12) to achieve the similar degree of convergence.

**Backbone Hydrogen Bonding**

Similar to the study by Honig and coworkers, not all the hydrogen bonds observed in the X-ray structures are consistently formed throughout the simulation, though our analysis gives more information about the spatial variations in the stability of these hydrogen bonds. The expected trend is that the hydrogen bonds buried away from the solvent tend to be more stable than the ones that are solvent exposed, and this trend is consistent between the explicit and implicit solvent models. The peaks in the lifetime (Fig. 13) and occupancy (Fig. 14) plots correspond to the hydrogen bonds that are either in the $\beta$-sheet, which forms the protein core, or on the inside of the helices that pack against the core. A few residues on the outside of the protein have relatively long lifetimes and high occupancies, such as Lys 70 and Ile 96, since their main-chain atoms are shielded from the solvent by the side-chains, similar to what was observed for lysine residues in polyalanine helices. There are also a few notable differences between the GBSW and explicit solvent results. For the UnP-Inact state, for example, there were no consistent differences between the occupancies (Fig. 14a), though a few residues in helical regions (e.g., Val 21 and Leu 25 in $\alpha_1$ and Lys 70 in $\alpha 3$) showed increased lifetimes in GBSW simulations (Fig. 13a), presumably because of the increased stabilization of hydrogen bonds in the GBSW model, as discussed earlier. For the P-Act state, by contrast, similar results were found for the lifetimes (Fig. 13d), but noticeably lower occupancies in the $\alpha 5$ helix were observed for the explicit solvent simulations runs (Fig. 14d). Since the active state crystal structure has an altered carboxyl terminal configuration due to defects in crystal packing, some instability was observed in this region (see RMSF in Fig. 6d). To illustrate what was taking place, representative snapshots of this helix are shown in Figure 15, where the protein was best-fit only with the $\alpha 5$ helix backbone for comparing the internal helical structure. The explicit solvent simulation resulted in a structure in which the later half of the $\alpha 5$ helix widened and exchanged hydrogen-bond contacts with other parts of the helix. The GBSW simulation, by contrast, maintained the $\alpha$-helix pattern; this observation once again supported the conjecture that the current parameterization of GBSW overestimates hydrogen-bonding interactions, which helps to stabilize secondary structures.

**Principal Component Analysis**

Quasiharmonic analysis probes the modes of motion sampled in MD simulations. Though similar in nature to normal mode analysis, quasiharmonic modes include the anharmonic contributions from the rugged energy landscape in the vicinity of the low energy conformations. The overlap matrices of quasiharmonic modes between a few unphosphorylated-inactive runs are displayed in Figure 16a and the best overlap is observed for the two GBSW runs. This point is better illustrated by the spanning coefficients, where the first 250 modes of one GBSW simulation span the modes of another GBSW simulation better than the comparison between two explicit solvent simulations or between an explicit solvent and a GBSW simulation. For all pair-wise comparisons, the lowest frequency modes, which dominate collective motions of the protein, tend to overlap well (spanning coefficients $\sim 0.9$); the higher frequency modes, which tend to be more localized in nature, overlap less well.

Since the low-frequency quasiharmonic modes dominate the structural fluctuations in MD simulations, the extent of conformational sampling for each protocol was essentially being compared. Therefore, it seems that the protein conformational space sampled with the GBSW model was more similar between simulations, compared with the space sampled by the explicit solvent simulations; the conformational space sampled by the GBSW and explicit solvent simulations were also not as similar. Whether these trends truthfully reflect better sampling in the GBSW simulations when compared
Comparison of Generalized Born Model with Explicit Solvent Simulations for CheY

Figure 13. Average lifetimes (see Computational Methods) of backbone hydrogen bonds. See Table 1 for notation. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

Conclusions

Because of the improved computational speed and increased efficiency for conformational sampling over explicit solvent simulations, implicit solvent simulations are potentially powerful in a wide range of biomolecular applications such as the prediction and refinement of structures.\textsuperscript{15,111–113} If it is the mechanistic detail of conformational transitions, rather than the final structure, are of interest, care must be taken to ensure that the implicit solvent model is appropriate for the problem and system at hand.

In the current work, we carried out systematic analyses on the results from simulations using a recently developed implicit solvent model, the GBSW method, for a signaling protein, CheY. The results were compared to explicit solvent simulations using the stochastic boundary condition. These studies were motivated by our interests in the mechanism of allosteric transitions in signaling proteins and molecular motors, for which identifying an efficient and reliable sampling technique is crucial. Moreover, most previous benchmark analysis of implicit solvent models focused on proteins with a single dominant conformation, and therefore, the question of whether popular implicit solvent models are likely appropriate for studying conformational transition problems in biomolecules remains unclear.

One unique feature of the current work is that the structural and dynamical properties analyzed include not only those localized to the active site region, but also throughout the protein; the signaling function of CheY also allowed us to examine these properties as a function of different conformational and phosphorylation states of the protein. The combination of these results offered a rather comprehensive and stringent test of the GBSW model in comparison to the explicit solvent simulations.
Overall, many properties were well reproduced by the GBSW simulations when compared with the explicit solvent calculations. These include not only rather global observables such as the Cα RMSF, but also fairly subtle quantities such as the side-chain order parameters for residues in the protein core (Ile 55, Met 85) and active site (Thr 87); the principal components are also fairly similar. These are quite impressive and encouraging findings regarding the applicability of the GBSW model. Nevertheless, several observations in the present work consistently pointed to the suggestion that the current parameterization of the GBSW model tends to overestimate hydrogen-bonding interactions involving both charged groups and (charge-neutral) backbone atoms. This deficiency led to overstabilization of certain secondary structural motifs; more importantly, qualitatively different behaviors compared with explicit solvent simulations were found for active site groups (Thr 87, Ala 88, the β4–α4 loop) in response to the change in the phosphorylation state of Asp 57, which are directly implicated in the function of CheY. Therefore, it seems that adjustment in the GBSW model is needed to make it more applicable to the mechanistic studies of conformational transition in biomolecules.

The simulation study of CheY also highlighted the value of carrying out both explicit solvent and implicit solvent simulations for mechanistic analysis. Although both methods have limitations and may lead to certain simulation artifacts, due to either limited sampling or inaccurate treatment of intermolecular interactions, combining the results is valuable for identifying robust mechanistic features. In the specific case of CheY, for example, although the behavior for the β4–α4 loop with the GBSW model was

---

While this was under review, we noted that such adjustments have been finished by Chen et al., who parameterized a new set of atomic radii for selected polar/charged groups in the CHARMM 22 force field and adjusted the protein backbone torsional energetics. This newly optimized GBSW model was found successful in reproducing the conformational equilibria for a range of helical and β-hairpin peptides and small proteins. It would be of tremendous interest to study CheY using this model in the near future and compare the results with the findings reported here.

---

Figure 14. Average occupancy (see Computational Methods) of backbone hydrogen bonds. See Table 1 for notation. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]
Comparison of Generalized Born Model with Explicit Solvent Simulations for CheY

Figure 15. Snapshot of the α5 helix at the end of simulation for the phosphorylated active state. (a) Explicit solvent simulation no. 1 after 7 ns; (b) GBSW simulation no. 1 after 9 ns. Both snapshots (colored by atom) were best-fit by backbone atoms to the active-state (1F4V)46 crystal structure, which was the starting structure for both simulations. Note that the backbone hydrogen bonding pattern was well-maintained throughout the GBSW simulation but altered at the end of the helix in the explicit solvent simulation. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

different from the explicit solvent simulations in the P-Inact and UnP-Act states, these results were in fact consistent with and helped to reinforce the idea that the β4→α4 loop configuration depends sensitively on the hydrogen bonds involving Thr 87 and Ala 88.

Finally, we emphasize that for the relatively small system such as CheY, the gain in computational speed, per se, over explicit solvent simulations can be rather modest because of the large cutoff used in the GBSW simulations. In the current work, the GBSW simulations were faster by only a factor of two than the stochastic boundary simulations; this is likely to increase as the system size increases. Moreover, as emphasized in Introduction, implicit solvent simulations may provide additional gain by more efficient sampling of the conformational space. Because of a skewed energy landscape, such as the overestimated hydrogen-bonding interactions in the current parameterization of GBSW, it is possible that not all processes are sped up with implicit solvent simulations; the convergence behavior for side-chain order parameters in explicit solvent and GBSW simulations discussed here is a good example. In fact, it is entirely possible that the relative time-scale for different processes is significantly altered with implicit solvent simulations. Whether this occurs often in realistic systems and has important mechanistic implications remains to be carefully analyzed.

Supplementary Material

Additional figures regarding the mainchain NH order parameters and comparison with side-chain methyl axial order parameter are included.
Figure 16. Overlap matrices (a–c) and spanning coefficients (d) for the unphosphorylated-inactive simulations with different solvent models. The scale for the overlap matrices is from 0.0 (white) to 1.0 (dark blue). [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

Acknowledgments

Q.C. is an Alfred P. Sloan Research Fellow. Computational resources from the National Center for Supercomputing Applications at the University of Illinois are greatly appreciated.

References
