*Protein Science* (1995), *4*:387-393. Cambridge University Press. Printed in the USA.


# Conservative and nonconservative mutations in proteins: Anomalous mutations in a transport receptor analyzed by free energy and quantum chemical calculations

WILLIAM R. CANNON,[1,3] JAMES M. BRIGGS,[1,4] JIAN SHEN,[1,5]
J. ANDREW McCAMMON,[1,4] AND FLORANTE A. QUIOCHO

[1] Department of Chemistry, University of Houston, Houston, Texas 77204-5641
[2] Howard Hughes Medical Institute, Departments of Biochemistry and Molecular Physiology and Biophysics, Baylor College of Medicine, Houston, Texas 77030

(RECEIVED August 10, 1994; ACCEPTED December 7, 1994)

**Abstract**

Experimental studies on a bacterial sulfate receptor have indicated anomalous relative binding affinities for the mutations $Ser_{130} \rightarrow$ Cys, $Ser_{130} \rightarrow$ Gly, and $Ser_{130} \rightarrow$ Ala. The loss of affinity for sulfate in the former mutation was previously attributed to a greater steric effect on the part of the Cys side chain relative to the Ser side chain, whereas the relatively small loss of binding affinity for the latter two mutations was attributed to the loss of a single hydrogen bond. In this report we present quantum chemical and statistical thermodynamic studies of these mutations. Qualitative results from these studies indicate that for the $Ser_{130} \rightarrow$ Cys mutation the large decrease in binding affinity is in part caused by steric effects, but also significantly by the differential work required to polarize the Cys thiol group relative to the Ser hydroxyl group. The Gly mutant cobinds a water molecule in the same location as the Ser side chain resulting in a relatively small decrease in binding affinity. Results for the Ala mutant are in disagreement with experimental results but are likely to be limited by insufficient sampling of configuration space due to physical constraints applied during the simulation.

**Keywords:** binding proteins; molecular recognition; polarization effects; protein engineering; simulation

Site-directed mutagenesis has become one of the most powerful and significant tools for molecular biology, biochemistry, and biotechnology. Protein engineering, binding studies of drugs and receptors, and mechanistic studies in enzymology and molecular biology all depend on correct interpretations of mutagenesis work. In fact, accurate interpretations can be the basis for considerable investments in time and money by researchers in academia and industry.

Several classes of mutations are considered to be relatively conservative, that is, substitution of one amino acid residue for a homologous one (such as isoleucine for leucine, aspartate for glutamate, or cysteine for serine) is not expected to perturb the protein significantly. Supposedly, conservative mutations that result in a loss of activity are therefore deserving of detailed analysis.

Sulfate binding protein (SBP), an initial receptor for sulfate translocation in *Salmonella typhimurium*, exhibits a 3,200-fold decrease in binding upon the substitution of $Ser_{130}$ with Cys, yet only a 100- and 15-fold decrease in binding when Ala or Gly are substituted, respectively (He & Quiocho, 1991). The loss of binding due to the former substitution has been attributed to a greater steric interaction on the part of the thiol group relative to the hydroxyl group by He and Quiocho (1991). Their conclusion was based in part on the fact that the binding gradually decreases from pH 5.0 to pH 8.0, corresponding to the ionization of the thiol as the pH increases.

The ability of the Ala and Gly mutants to bind the dianion is also not well understood. Previously, the relatively small decrease in binding has been attributed by He and Quiocho (1991) to the loss of one protein-sulfate hydrogen bond, with the apolar $\beta$-methyl side chain of Ala causing a further sevenfold decrease in binding relative to Gly, presumably by steric interaction.

Reprint requests to: J. Andrew McCammon, Department of Chemistry and Biochemistry, University of California, La Jolla, California 92093-0365; e-mail: jmccammon@ucsd.edu.
[3] Present address: 152 Davey Laboratory, Department of Chemistry, Penn State University, University Park, Pennsylvania 16801.
[4] Present address: Department of Chemistry and Biochemistry, and Department of Pharmacology, University of California at San Diego, La Jolla, California 92093-0365.
[5] Present address: Sterling-Winthrop, Inc., P.O. Box 5000, Collegeville, Pennsylvania 19426-0900.

SBP is a bilobate protein that binds the sulfate dianion using only dipolar interactions and no salt bridges. Five backbone amide hydrogens, the NH of the $Trp_{192}$ side chain, and the hydroxyl hydrogen of $Ser_{130}$ participate in binding the dianion. In all, three $\alpha$-helices terminate their electropositive ends at the binding site.

In this paper, we present a computational study of the binding of mutant SBP relative to the native protein. Ab initio calculations on $CH_3OH \cdot SO_4^{2-}$ and $CH_3SH \cdot SO_4^{2-}$ with an appropriate basis set are used to develop a description of the differential interactions between $Ser \cdot SO_4^{2-}$ and $Cys \cdot SO_4^{2-}$ in the protein. Polarization effects are taken into account explicitly in these calculations, and the work required to polarize the amino acid side chains is determined from the partial charges found in the in vacuo and sulfate-complexed species. Subsequently, relative free energies of binding are determined for the $Ser_{130} \rightarrow Cys$, $Ser_{130} \rightarrow Ala$, and $Ser_{130} \rightarrow Gly$ mutants by computer simulations and statistical mechanical analysis. The results are used to aid interpretation of the experimental data of He and Quiocho.

## Results and discussion

Quantum chemical calculations on sulfate and the side-chain analogues $CH_3OH$ and $CH_3SH$ were used to characterize binding to $SO_4^{2-}$ as far as binding energies, intermolecular distances (H-bond length), and charge distribution. Ab initio calculations for the configurations shown in Figure 1 result in binding energies of $-82.4$ and $-69.0$ kJ mol$^{-1}$ at the HF/6-31+G(d)//6-31+G(d) level and $-81.2$ and $-61.9$ kJ mol$^{-1}$ at the HF/6-31+G(d,p)//6-31+G(d) level for $CH_3OH \cdot SO_4^{2-}$ and $CH_3SH \cdot SO_4^{2-}$, respectively. The calculations at the HF/6-31+G(d,p)//6-31+G(d) level also included the counterpoise estimate of basis set superposition error (Boys & Bernardi, 1970), which attempts to correct for errors introduced by the use of finite basis sets. Partial charges from fitting the electrostatic potential to point charges are shown in Figure 1. Charges are summarized in Table 1 for environments without any external field (in vacuo) and in the electric field of the $SO_4^{2-}$ anion, along with the OPLS (Jorgensen & Tirado-Rives, 1988) charges used in the simulations for comparison. The ab initio charges do not necessarily sum to zero due to round-off error and a small charge transfer to the anion. The OPLS parameters utilize a united atom methyl group. As can be seen, polarization occurs in both methanethiol and methanol but is more significant in the former. The work required create the induced dipole can be determined from $U_{pol} = 1/2\mu \cdot \mu/\alpha$ (Berendsen et al., 1987) where $\mu$ is the induced dipole moment estimated from partial charges of the in vacuo and $SO_4^{2-}$-complexed species and $\alpha$ is the molecular polarizability. The polarizability $\alpha$ is $3.26 \cdot 10^{-3}$ nm$^3$ and $5.62 \cdot 10^{-3}$ nm$^3$ for $CH_3OH$ and $CH_3SH$, respectively. The former value is known from experimentation and the latter value was calculated by the method of Miller and Savchik (1979). Creation of the induced dipole in methanol requires 5.44 kJ mol$^{-1}$ of work, whereas the corresponding quantity for the thiol is 16.53 kJ mol$^{-1}$. This significant difference is at least partially responsible for the decreased ability of $CH_3SH$ to bind $SO_4^{2-}$ relative to $CH_3OH$ in vacuo. These polarization effects could be expected to be present in the water-$SBP \cdot SO_4^{2-}$ system and are not likely to cancel out in the unliganded water-SBP system. First, the water-protein system involves dipole-dipole interactions as opposed to charge-dipole interactions. Although the weaker water dipoles can also polarize $CH_3SH$ and $CH_3OH$, the average configuration of waters around these solutes will not be as directional as the liganded protein environment. The dynamic behavior of the $SO_4^{2-}$-protein

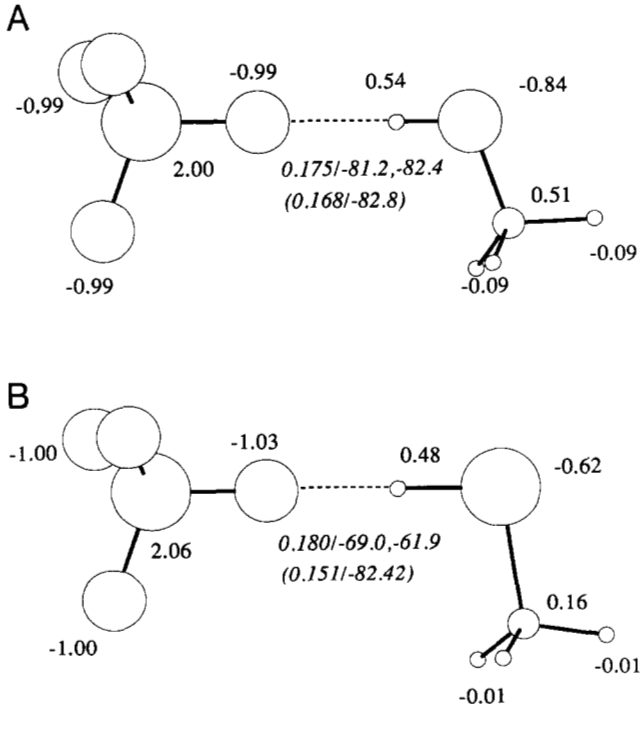


**Fig. 1.** Ab initio charges, optimized distances, and energies of complexes with sulfate with molecular mechanics values shown in parentheses. **A:** $CH_3OH \cdot SO_4^{-2}$. **B:** $CH_3SH \cdot SO_4^{-2}$. Dotted lines indicate hydrogen bonds and numbers below the dotted lines indicate heavy atom to H distance/HF/6-31+G(d) energy, HF/6-31+G(d,p) BSSE corrected energy. Distances are in nm and energies in kJ mol$^{-1}$.

**Table 1.** *Charges from ab initio calculations for free and complexed species of methanol and methanethiol*[a]

| | | Charges | | |
|---|---|---|---|---|
| | | Ab initio | | |
| | Atom | Free | Complex | OPLS |
| $CH_3SH$ | S | −0.38 | −0.62 | −0.45 |
| | H(S) | 0.21 | 0.48 | 0.27 |
| | C | 0.06 | 0.16 | 0.18 |
| | H(C) | 0.09 | −0.01 | |
| | H(C) | 0.01 | −0.01 | |
| | H(C) | 0.01 | −0.01 | |
| $CH_3OH$ | O | −0.75 | −0.84 | −0.70 |
| | H(O) | 0.44 | 0.54 | 0.435 |
| | C | 0.39 | 0.51 | 0.265 |
| | H(C) | 0.01 | −0.09 | |
| | H(C) | −0.05 | −0.09 | |
| | H(C) | −0.05 | −0.09 | |

[a] OPLS charges shown for comparison.

system near the active site is expected to be limited based on crystallographic temperature factors below 0.08 $nm^2$ for atoms in this region.

For comparison, binding studies were done with the molecular mechanics potential functions in which $SO_4^{2-}$ (Cannon et al., 1994) and OPLS potentials were used. Energies and intermolecular distances were similar for $CH_3OH \cdot SO_4^{2-}$ in the ab initio and molecular mechanics calculations as shown in Figure 1. However, also as shown in Figure 1, the molecular mechanics potential functions for $CH_3SH \cdot SO_4^{2-}$ resulted in a binding energy several kJ $mol^{-1}$ too high and a thiol-sulfate hydrogen bond that is about 0.03 nm too short in comparison with ab initio results. Development of the OPLS functions for sulfhydryl groups involved binding studies with water and cations, dimerization, and liquid simulations. No binding studies of the sulfhydryl compounds with anions were reported. Many properties of alcohols (Jorgensen, 1986b) and sulfhydryl compounds (Jorgensen, 1986a) can be reproduced quite well with the OPLS pair potentials for $CH_3SH$ and $CH_3OH$. However, these parameters may require further development for use with anions. The bond distance for the heavy atom hydrogen is only ca. 0.095 nm for alcohols, whereas for sulfhydryl compounds, the distance is typically 0.134 nm. Even though the atomic diameter for sulfur ($\sigma = 0.355$ nm) is larger than for oxygen ($\sigma = 0.307$ nm), the actual potential should also be softer to account for the greater polarizability of sulfur. This may necessitate that explicit Lennard-Jones terms be used for the hydrogen if pair potentials for use with anions are to be developed.

Free energy simulations were carried out to determine binding affinities of the $Cys_{130}$, $Gly_{130}$, and $Ala_{130}$ mutants for the $SO_4^{2-}$ ligand relative to the wild-type protein. The starting points for these calculations were the crystallographic structure of the closed, liganded SBP and an open, unliganded structure based on structural homology with binding proteins of the same family (see Methods section and Quiocho [1990]). Results for the free energy simulations are shown in Table 2 and the convergence of the simulations as a function of simulation length is illustrated in Figure 2. The $Ser_{130} \rightarrow Cys$ mutation results in a free energy change for binding of 11.4 kJ $mol^{-1}$ in comparison with the experimental results of 20.1 kJ $mol^{-1}$. Although the calculation is qualitatively consistent with the experimental result, the agreement could be expected to improve with refinement of the OPLS Lennard-Jones parameters as described above. In particular, inclusion of a repulsive center on the thiol H would tend to increase the thiol-sulfate separation. The side chain of residue 130 exists almost exclusively in the hydrogen bonded configuration shown in Figure 3 for both the Ser and Cys endpoints and throughout the simulation. The 11.4-kJ $mol^{-1}$ free energy difference in this case reflects in part the work needed to create the methanethiol geometry from the methyl alcohol geometry in the environment of the complexed protein relative to the solvated open form. This large free energy change for what seems a modest change in geometry reflects the highly structured nature of this binding site. Part of this work is required to push $SO_4^{2-}$ away from other hydrogen bonding partners, including the amide hydrogens being donated by residues 131 and 132, as illustrated in Figure 3. A comparison of the side-chain geometries is shown in Table 3. However, as mentioned above and

**Table 2.** *Results from free energy simulations in kJ $mol^{-1}$*

| | Relative free energy | |
|---|---|---|
| Mutation | TI | Exp |
| Ser 130 Gly | 5.3 ± 12 | 6.7 |
| Ser 130 Cys | 11.4 ± 2 | 20.1 |
| Ser 130 Ala | 29.8 ± 6 | 11.3 |
| Ala 130 Gly | −24.5 ± 6 | |

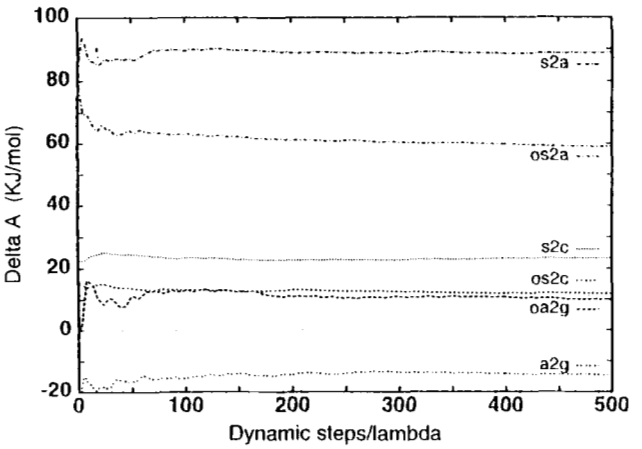


**Fig. 2.** Convergence of free energy simulations following 500 steps of equilibration in which the dependence of the free energy difference on the number of dynamic steps collected in the ensemble at each increment of λ during the thermodynamic integration is shown. The $Ser_{130} \rightarrow Cys_{130}$ simulation is indicated by s2c, $Ala_{130} \rightarrow Gly_{130}$ simulation is indicated by a2g, $Ser_{130} \rightarrow Ala_{130}$ simulation is indicated by s2a, and an "o" preceding the symbolic label refers to the open-form unliganded simulation.

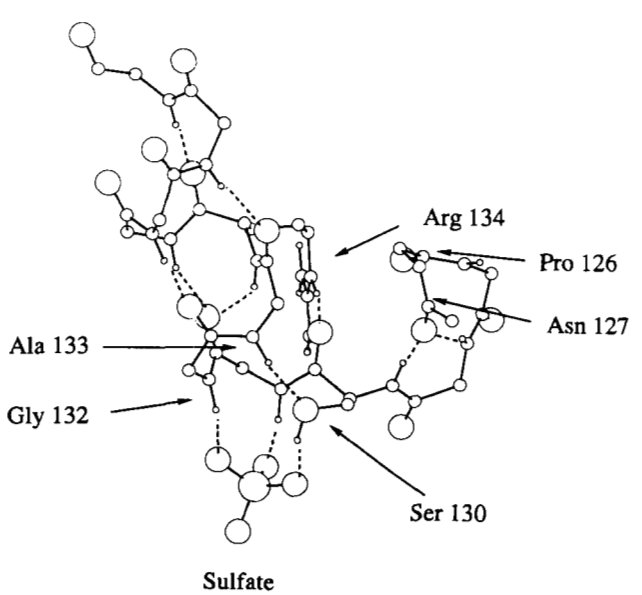


**Fig. 3.** Partial structure of the sulfate binding site showing the hydrogen bonding network involved in binding.

**Table 3.** *Geometric parameters for the Ser and Cys side chains, where X denotes either oxygen or sulfur*[a]

| | Methyl alcohol | Methyl thiol |
|---|---|---|
| $r_{H-X}$ | 0.095 | 0.133 |
| $r_{X-C}$ | 0.140 | 0.182 |
| $\theta_{HXC}$ | 110.38 | 97.90 |

[a] Distances are in nanometers.

not accounted for in the simulations, the work required to polarize the thiol also makes a significant contribution to the net preference for binding to the wild-type protein.

The result for the $Ser_{130} \rightarrow$ Gly mutation of 5.3 kJ $mol^{-1}$ agrees well with the experimental value of 6.7 kJ $mol^{-1}$. The small free energy change can be directly attributed to the movement of a water into the space previously occupied by the serine hydroxyl group during the simulation, as shown in Figure 4. The water molecule binds in nearly the identical manner as the Ser hydroxyl group, accepting a hydrogen bond from the backbone amide of $Ala_{133}$ and donating one to $SO_4^{2-}$. The water had previously been located 0.5 nm away from the $SO_4^{2-}$ and was the closest water to the binding site. It is important to note that the water was not grown in by thermodynamic integration, but moves into the binding site as space is created by the removal of the methoxy group of serine. Because free energy simulations are based on the principles of equilibrium statistical mechanics care must be taken to ensure that nonequilibrium processes do not occur. The rate at which this water can move is limited by the diffusion constant of water, which is $2.4 \times 10^{-5}$ $cm^2$ $s^{-1}$ for real water, $4.3 \times 10^{-5}$ $cm^2$ $s^{-1}$ for normal mass SPC bulk water, and approximately $3.0 \times 10^{-5}$ $cm^2$ $s^{-1}$ for bulk water with a hydrogen mass of 10 amu. Using these values, the time required for reversible movement into the active site can be estimated to be 52, 29, and 42 ps, respectively, which is well within the simulation time of 94.5 ps. The system is closed and no corrections to the free energy change are required to account for the cobinding of the water with the sulfate dianion. Of course, in the physical binding process this water is likely a first solvation shell water that is not required to be stripped away in order for $SO_4^{2-}$ to bind. The statistical error estimate of 12 kJ $mol^{-1}$ associated with this free energy change results from the combination of the $Ser_{130} \rightarrow$ Ala simulation (below) with an $Ala_{130} \rightarrow$ Gly simulation (Table 2), both of which have estimated statistical errors of 6 kJ $mol^{-1}$. The magnitude of this error is likely to be due to the creation of a void as the methoxy group is withdrawn into the van der Waals region of the backbone $\alpha$-carbon. Although the movement of the water into the binding site is significant for interpretation of the experimental findings, the large statistical error limits the simulation free energy value to be of a qualitative nature only.

The result for the $Ser_{130} \rightarrow$ Ala mutation of 29.7 kJ $mol^{-1}$ is in significant quantitative disagreement with the experimentally determined value of 11.3 kJ $mol^{-1}$. Although it is possible that during the simulation the nearby waters do not have the time to diffuse into close proximity to the anion or cannot overcome local barriers to solvate the anion as seen in the previous simulation, it may be more significant that the methyl side chain of Ala and a water molecule occupy a greater volume than the Ser methoxy side chain. The protein may have to open up slightly to accommodate this water. However, as only the atoms within 1.8 nm of $Ser_{130}$ are allowed to move (see Methods), the simulation does not allow for large-scale breathing motions that have been demonstrated by kinetic studies to be necessary for binding processes (Jacobson et al., 1991, 1992) and that could facilitate the entry and binding of a water molecule in the Ala mutant. To address this issue we continued the liganded $Ser_{130} \rightarrow$ Ala simulation for 1,000 equilibration steps and 2,000 data gathering steps for a total simulation time of 283.5 ps. Although the free energy for this leg of the thermodynamic cycle dropped by 2 kJ $mol^{-1}$, nearby water did not move into the void created by the removal of the Ser hydroxy group. As shown in Figure 5, the free energy convergence remained relatively flat during the extended simulation, indicating that no relaxation processes occurred on this time scale that were unevenly sampled and were relevant to the binding of $SO_4^{2-}$. However, it is important to bear in mind that the relaxation time for the H-bond network in the $SO_4^{2-}$·SBP complex is unusually long and likely to be on the order of a second (Jacobson et al., 1991), and therefore the complete sampling of phase space is not possible with current computers and algorithms. Previous studies (Wade et al., 1991) with this protein have indicated that the cost of inserting a water into a cavity in the protein is in the range of 0.8 kJ $mol^{-1}$. However, the cavity used in the previous study was chosen to be one of sufficient volume to accommodate a water molecule,

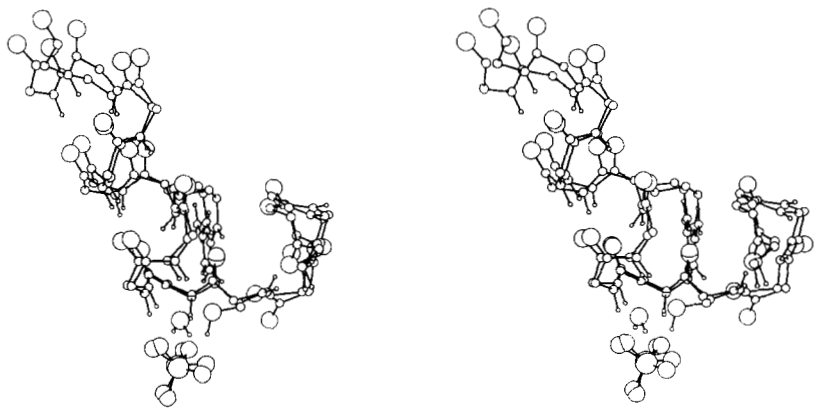

**Fig. 4.** Partial stereo structures of SBP binding site showing native structure overlayed on Gly mutant where water occupies space created by removal of serine methoxy group.

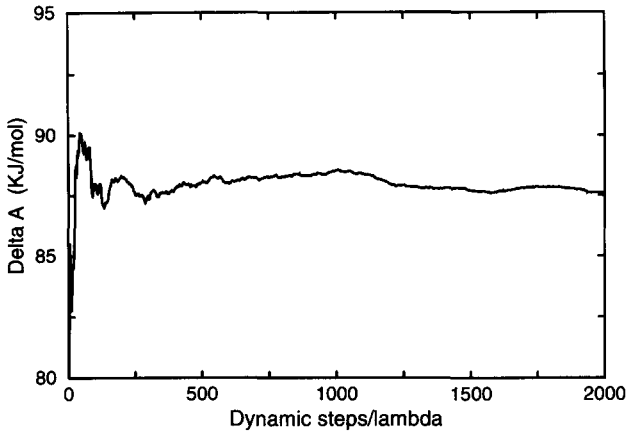


**Fig. 5.** Convergence of data gathering steps following 1,000 steps of equilibration for the liganded $Ser_{130} \rightarrow Ala_{130}$ simulation.

and it was located far away from the sulfate binding site to decrease any influence of the ion on insertion of the water. The cavity in this case is literally next to the sulfate ion and is characterized by the highly ordered hydrogen bonding network shown in Figure 3, which may limit the conformational freedom of the protein to adjust accordingly. The energetic cost of inserting a water in this location may differ significantly from what was found in the previous study.

A potential source of error in all three systems is the use of the homology-built model for the unliganded, open form of the binding protein (see Methods) even though other binding proteins in this family exhibit high structural homology for both unliganded and liganded forms (Sharff et al., 1992; Olah et al., 1993). Although the detailed structure of the unliganded protein is not known, it is reasonable to infer from the crystallographic studies on the related transport proteins that the binding site residues are exposed to the solvent. Given this, the exact configuration of the protein in the open, unliganded form may not be as important as the local environment of the residue of interest ($Ser_{130}$), which is presumably partially solvated. It has been shown in previous studies (Jorgensen & Briggs, 1989) that the OPLS parameters used here give the correct free energy change for the solvation of methanethiol relative to methanol.

## Conclusion

The goal of this study has been to aid in the interpretation of anomalous binding affinities in a transport receptor. The $Ser_{130} \rightarrow$ Cys mutation in SBP results in a significant change in sulfate-binding affinity despite a relatively conservative change in side-chain geometry. Although the simulation results in a relative free energy change in binding of 11.4 kJ $mol^{-1}$, the calculated value would likely be closer to the experimental value of 20.1 kJ $mol^{-1}$ if the molecular mechanics model were refined to maintain the relatively large thiol-sulfate separation seen in the ab initio calculations. The simulations and ab initio calculations taken together suggest that the large free energy change observed experimentally results in part from steric effects, but also from the greater amount of work required to polarize the Cys thiol group relative to the Ser hydroxyl group in the proximity of $SO_4^{-2}$. For the $Ser_{130} \rightarrow$ Gly mutant, the energetic cost of losing a hydrogen bond to the Ser hydroxyl group is partly compensated by a water molecule that binds with the anion. The same phenomenon likely occurs for the $Ser_{130} \rightarrow$ Ala mutant, but this was not seen during the simulation and may be due to the constraints applied during the simulation.

Simulations involving high charge densities present special problems. Specifically, because the forces between charges can be quite large, exploration of phase space is difficult due to the deep potential wells and long relaxation times that can lead to sampling errors. Of the seven protein dipole–sulfate interactions, only the $Ser_{130}$ hydrogen bond has any significant rotational freedom; the other hydrogen bonds come from helix-backbone amide interactions or the $Trp_{192}$ NH interaction. It is important to note that a fairly large reorganization of the binding site in which the lost serine hydroxyl hydrogen bond is replaced by a protein backbone amide hydrogen bond from a disrupted helix IV cannot be ruled out because the hydrogen bond network relaxation time would be longer than the simulation time. Also, the use of the homology built model for the unliganded, open form structure may be inadequate in spite of the high structural conservation seen for the other unliganded proteins in this family. However, it can be inferred that the $Ser_{130} \rightarrow$ Ala mutant, like the $Ser_{130} \rightarrow$ Gly mutant, co-binds a water molecule along with the sulfate by opening up the binding pocket slightly. Subsequent computational studies of this system and others may require refinement of the Lennard-Jones parameterization as discussed above and perhaps explicit polarizability of key residues and mobility of the full protein in order to allow large-scale breathing motions.

## Methods

### *Ab initio calculations*

Ab initio quantum chemical calculations were carried out to study the binding of $SO_4^{2-}$ to the Cys and Ser side chain analogues $CH_3SH$ and $CH_3OH$. The 6-31+G(d) basis set, which contains polarization functions for non-hydrogen atoms, was employed for geometry optimization of the intermolecular distance with a spin-restricted, closed-shell wave function using Gaussian 92 (Frisch et al., 1992). The optimized HF/6-31+G(d) geometries were then used in single point HF/6-31+G(d,p), which includes polarization functions on both hydrogen and non-hydrogen atoms. This latter calculation also included a counterpoise correction (Boys & Bernardi, 1970) for the basis set superposition error. Although electron correlation is important for many systems, previous work (Cannon et al., 1994) done on sulfate complexes suggests that errors due to basis set superposition errors are more significant. The $CH_3OH \cdot SO_4^{2-}$ and $CH_3SH \cdot SO_4^{2-}$ complexes were held in a monodentate binding mode in which $CH_3OH$ and $CH_3SH$ approach the anion such that the interacting hydrogen and heteroatom are colinear with a sulfate oxygen and sulfur, and the methyl carbon is coplanar with the same sulfate oxygen, sulfur, and a second sulfate oxygen as shown in Figure 1. Only the intermolecular distances between the species were optimized. The tetrahedral sulfate geometry utilized here and in the simulations used a bond length of 0.14898 nm (Cannon et al., 1994). The geometries for $CH_3OH$ and $CH_3SH$ were from previous optimizations at the HF/6-31+G(d) level (Jorgensen & Briggs, 1989). Partial charges on the atoms were obtained for the complex from fitting the quantum

$$\begin{array}{ccccc} & & & \xrightarrow{\Delta A_1} & \\ \mathrm{SBP} & + & \mathrm{SO_4^{-2}} & \longrightarrow & \mathrm{SBP{\cdot}SO_4^{-2}} \\ & \Delta A_3 & \downarrow & & \downarrow \ \Delta A_4 \\ \mathrm{SBP\prime} & + & \mathrm{SO_4^{-2}} & \xrightarrow[\Delta A_2]{} & \mathrm{SBP\prime{\cdot}SO_4^{-2}} \end{array}$$

$$\Delta\Delta A = \Delta A_2 - \Delta A_1 = \Delta A_4 - \Delta A_3 \qquad (1)$$

$$\Delta A = \int \left\langle \left( \frac{\partial H}{\partial \lambda} \right)_{T,V} \right\rangle_{\lambda} d\lambda \qquad (2)$$

**Fig. 6.** Equations 1 and 2. Equation 1 denotes the thermodynamic cycle employed and Equation 2 the thermodynamic integration equation implemented for determination of free energy differences.

mechanical electrostatic potential to a point charge model with a least-squares fit (Breneman & Wiberg, 1990).

*Free energy simulations*

A thermodynamic cycle (Tembe & McCammon, 1984) is employed to obtain the free energies of binding of the various mutants relative to the native protein in solution, as shown in Equation 1 (Fig. 6). Computationally, the native SBP is slowly changed into one of the mutant proteins (SBP′) and the free energy differences $\Delta A_3$ and $\Delta A_4$ are calculated.

Free energy differences can be obtained via computer simulations by the use of either the thermodynamic integration or thermodynamic perturbation method (Straatsma & McCammon, 1992). The nature of the Hamiltonian determines the state of the system, and two states can be coupled by the variable λ. Using thermodynamic integration, the free energy difference between two states at constant volume and temperature is obtained from Equation 2 (Fig. 6), where $H$ is the system Hamiltonian, λ is the state coupling variable, and the brackets indicate an ensemble average. In the thermodynamic integration method implemented here (Straatsma & McCammon, 1991; Mark et al., 1994), the average of the derivative of the Hamiltonian is computed at discrete values of the integration variable λ. A quadrature over these values gives the difference in the Helmholtz free energy of the two systems of interest. Because of the statistical nature of the method, error estimates can also be obtained (Straatsma & McCammon, 1991).

The OPLS (Jorgensen & Tirado-Rives, 1988) nonbonded potential parameters were used with GROMOS (van Gunsteren & Berendsen, 1987) bonded parameters in the simulation. In order to utilize the OPLS parameters, 1,4 electrostatic and Lennard-Jones interactions are scaled by 1/2 and 1/8, respectively, except for 1,4 Lennard-Jones interactions between terminal and side-chain polar species, which are not scaled. As noted previously (Smith et al., 1991) and also observed in our work, this is necessary in order to have proper energetics between these charged groups and the polar amino hydrogens. Sulfate parameters have been developed previously (Cannon et al., 1994). Calculations were done with ARGOS software (Straatsma & McCammon, 1990).

The starting structure for the liganded SBP simulation is the refined 1.7-Å crystal structure, 1SBP in the Brookhaven Protein Data Bank (J.S. Sack & F.A. Quiocho, in prep.), with the sulfate dianion sequestered in the binding pocket. The starting structure for the unliganded, open form SBP simulation was built from the structurally homologous leucine-isoleucine-valine binding protein, in which the ligand-binding cleft of the protein is open. The SBP structure was overlayed on the homologous leucine-isoleucine-valine binding protein structure, and the three dihedral angles that serve as a hinge in the bilobal protein were adjusted so that the protein backbone of SBP overlays the backbone of the leucine-isoleucine-valine binding protein. Structural homology of SBP and leucine-isoleucine-valine binding protein as well as other binding proteins has been inferred from crystallographic studies (Sack et al., 1989; Quiocho, 1990), and the relationship of cleft opening and ligand binding has been supported by kinetic studies (Jacobson et al., 1991, 1992).

Both the liganded and open, unliganded structures were solvated with SPC water in a 2.4-nm sphere centered on $\mathrm{Ser_{130}}$. No buried water molecules were generated by this procedure. The solvent molecules out to 1.8 nm from $\mathrm{Ser_{130}}$ were relaxed with 500 steps of steepest descent energy minimization. Molecular dynamics simulations using bond constraints as implemented in the SHAKE algorithm and time steps of 2.5 fs were performed on the water with the protein fixed for 5.0 ps at each of the temperatures 20, 70, 120, and 200 K, using coupling to an external temperature bath (Berendsen et al., 1984) with a system relaxation time of 0.1 ps. Temperature, potential energy, and total energy were used as indicators of equilibration of the system. Nonbonded cutoffs of 0.8 and 1.2 nm were used for short-range and long-range interactions, respectively.

Next, 500 steps of steepest descent energy minimization were performed on the protein only, followed by 5 ps of molecular dynamics at 20, 70, 120, and 200 K again using a time step of 2.5 fs with coupling to an external temperature bath. Protein and water were then allowed to move simultaneously for 10 ps at 200 and at 293 K.

At this point, hydrogens were assigned a mass of 10 amu and the time step was increased to 4.5 fs (Pomes & McCammon, 1990). Final equilibration was carried out at 293 K for 30 ps. Overlaying of the equilibrated structure with the initial structure from diffraction studies showed a high degree of similarity.

Thermodynamic integration for each of the six simulations was carried out in 21 increments of λ with 500 equilibration steps followed by 500 data gathering steps for a total of 21(500 + 500) = 21,000 dynamics steps of 4.5 fs each, or 94.5 ps total simulation time. Again, the system was coupled to an external temperature bath with a relaxation time of 0.4 ps.

## Acknowledgments

This work was supported in part by the NSF, the NIH, the R.A. Welch Foundation, the W.M. Keck Foundation, and the Metacenter program of the NSF Supercomputer Centers. W.R.C. is a predoctoral trainee in the NIH Houston Area Training Program in Molecular Biophysics.

## References


Berendsen HJC, Grigera JR, Straatsma TP. 1987. The missing term in effective pair potentials. *J Phys Chem 91*:6269–6271.

Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR. 1984. Molecular dynamics with coupling to an external bath. *J Chem Phys 81*:3684–3690.

Boys SF, Bernardi F. 1970. The calculation of small molecular interactions by the differences of separate total energies. Some procedures with reduced errors. *Mol Phys 19*:553–566.

Breneman C, Wiberg K. 1990. Determining atom-centered monopoles from molecular electrostatic potentials—The need for high sampling density in formamide conformational analysis. *J Comput Chem 11*:361–373.

Cannon WR, Pettitt BM, McCammon JA. 1994. Sulfate anion in water: Model structural, thermodynamic and dynamic properties. *J Phys Chem 98*:6225–6230.

Frisch M, Head-Gordon M, Trucks G, Foresman J, Schlegel H, Raghavachari K, Robb M, Binkley J, Gonzalez C, Defrees D, Fox D, Whiteside R, Seeger R, Melius C, Baker J, Martin R, Kahn L, Stewart J, Topiol S, Pople J. 1992. *Gaussian 92*. Pittsburgh, Pennsylvania: Gaussian, Inc.

He JJ, Quiocho FA. 1991. A nonconservative serine to cysteine mutation in the sulfate-binding protein, a transport receptor. *Science 251*:1479–1481.

Jacobson BL, He JJ, Lemon DD, Quiocho FA. 1992. Interdomain salt bridges modulate ligand-induced domain motion of the sulfate receptor protein for active transport. *J Mol Biol 223*:27–30.

Jacobson BL, He JJ, Vermersch PS, Lemon DD, Quiocho FA. 1991. Engineered interdomain disulfide in the periplasmic receptor for sulfate transport reduces flexibility. *J Biol Chem 266*:5220–5225.

Jorgensen WJ. 1986a. Intermolecular potential functions and Monte Carlo simulations for liquid sulfur compounds. *J Phys Chem 90*:6379–6388.

Jorgensen WJ. 1986b. Optimized intermolecular potential functions for liquid alcohols. *J Phys Chem 90*:1276–1284.

Jorgensen WL, Briggs JM. 1989. A priori $pK_a$ calculations and the hydration of organic anions. *J Am Chem Soc 111*:4190–4197.

Jorgensen WJ, Tirado-Rives J. 1988. The OPLS potential functions for proteins. Energy minimizations for crystals of cyclic peptides and crambin. *J Am Chem Soc 110*:1657–1666.

Mark AE, van Helden SP, Smith PE, Janssen LHM, van Gunsteren WF. 1994. Convergence properties of free energy calculations: $\alpha$-Cyclodextrin complexes as a case study. *J Am Chem Soc 116*:6293–6302.

Miller KJ, Savchik JA. 1979. A new empirical method to calculate average molecular polarizabilities. *J Am Chem Soc 101*:7206–7213.

Olah GA, Trakhanov S, Trewhella J, Quiocho FA. 1993. Leucine/isoleucine/valine-binding protein contracts upon binding of ligand. *J Biol Chem 268*:16241–16247.

Pomes R, McCammon JA. 1990. Mass and step length optimization for the calculation of equilibrium properties by molecular dynamics simulation. *Chem Phys Lett 166*:425–428.

Quiocho FA. 1990. Atomic structures of periplasmic binding proteins and the high-affinity active transport systems in bacteria. *Phil Trans R Soc Lond B 326*:341–352.

Sack JS, Saper MA, Quiocho FA. 1989. Periplasmic binding protein structure and function. *J Mol Biol 206*:171–191.

Sharff AJ, Rodseth LE, Spurlino JC, Quiocho FA. 1992. Crystallographic evidence of a large ligand-induced hinge-twist motion between the two domains of the maltodextrin binding protein involved in active transport and chemotaxis. *Biochemistry 31*:10657–10663.

Smith PE, Dang LX, Pettitt B 1991. Simulation of the structure and dynamics of the bispenicillamine enkephalin zwitterion. *J Am Chem Soc 113*:67–73.

Straatsma TP, McCammon JA. 1990. ARGOS, a general vectorized molecular dynamics program. *J Comput Chem 11*:943–951.

Straatsma TP, McCammon JA. 1991. Multiconfiguration thermodynamic integration. *J Chem Phys 95*:1175–1188.

Straatsma TP, McCammon JA. 1992. Computational alchemy. *Annu Rev Phys Chem 43*:407–435.

Tembe BL, McCammon JA. 1984. Ligand-receptor interactions. *Comput Chem 8*:281–283.

van Gunsteren WF, Berendsen HJC. 1987. *Gronigen molecular simulation (GROMOS) library manual*. Gronigen: CRC Press, Biomos.

Wade RC, Mazor MH, McCammon JA, Quiocho FA. 1991. A molecular dynamics study of thermodynamic and structural aspects of the hydration of cavities in proteins. *Biopolymers 31*:919–931.