JCTC Journal of Chemical Theory and Computation

Ligand Entropy in Gas-Phase, Upon Solvation and Protein Complexation. Fast Estimation with Quasi-Newton Hessian

S. Wlodek,* A. G. Skillman, and A. Nicholls

OpenEye Scientific Software Incorporated, 9 Bisbee Court, Suite D, Santa Fe, New Mexico 87508

Abstract: A method of rapid entropy estimation for small molecules in vacuum, solution, and inside a protein receptor is proposed. We show that the Hessian matrix of second derivatives built by a quasi-Newton optimizer during geometry optimization of a molecule with a classical molecular potential in these three environments can be used to predict vibrational entropies. We also show that a simple analytical solvation model allows for no less accurate entropy estimation of molecules in solution than a physically rigorous but computationally more expensive model based on Poisson's equation. Our work also suggests that scaled particle theory more precisely estimates the hydrophobic part of solvation entropy than the using a simple surface area term.

Introduction

The estimation of ligand entropy in different environments (vacuum, solution, and protein receptors) is essential for predicting the free energy of ligand transfer between them. While the prediction of entropy of the gas-phase compounds under low and moderate pressures can be achieved using basic statistical thermodynamics expressions for ideal gases,[1] provided that a set of a compound's normal frequencies is given, estimation of that state function in condensed phases is more difficult. The configurational part of entropy, S_c, is given by

$$S_c = -R \int P(\mathbf{r})\ln P(\mathbf{r})d\mathbf{r} \qquad (1)$$

where R is the gas constant and $P(\mathbf{r})$ is the probability density of the configuration given by coordinates $\mathbf{r}$, which is often determined from MD simulation of the system where $P(\mathbf{r})$ is derived from the accumulated trajectory[2] or, assuming that $P(\mathbf{r})$ is a multivariate Gaussian, from the quasiharmonic analysis of the diagonalization of the covariance matrix σ of the coordinate fluctuations:[3,4]

$$\sigma_{ij} = (r_i - \langle r_i \rangle)(r_j - \langle r_j \rangle) \qquad (2)$$

A similar method in which entropy is expressed as a function of coordinate variance, $\langle \Delta r^2 \rangle$, derived from MD simulation was proposed by Schlitter.[5] All such MD-based methods suffer from the large central processing unit (CPU) times necessary to properly cover phase space, and no matter how long a trajectory is generated, it is always incomplete. Some alternative methods apply corrected versions of the ideal gas-type entropy expressions, particularly to account for finite molecular volume in the translational part of entropy.[6–8] Although such a correction does indeed eliminate the overestimation of the entropy of a compound in solution upon the usage of ideal gas-type translational expression, in our opinion, it is not well founded because such a reduction is totally accounted for by the entropy effects of solvation phenomena and can be quantitatively described within an adopted solvation model, as, for example, in the recent work of Graziano.[9] In such a case, the Sackur−Tetrode equation can still be used to estimate the translational part of solute entropy.

These approaches rarely address the issue of conformational entropy. Conformational entropy as a part of configurational ligand entropy was evaluated by Gilson and colleagues[10,11] and recently applied to protein−ligand binding;[12] however their "mining minima" algorithm requires tens of hours on a commodity computer.

* Corresponding author. E-mail: stan@eyesopen.com.

None of the above approaches are suitable when the evaluation of entropy has to be done for a large number of ligands, for example, during drug design research. There is a need, therefore, for a rapid and reliable method of ligand entropy evaluation.

There are many factors that make the rapid and reliable estimation of ligand entropy in the condensed phase, either in solution or inside a protein receptor, a difficult task. One factor is the conformational diversity of a ligand in solution. Another is the necessity to include external forces acting on ligands in these environments: solvation forces in the case of solution ligands and protein–ligand intermolecular forces for protein-bound ligands. In the latter case, the shape of the protein–ligand potential well limits the motion of the ligand and, therefore, modifies its entropy. Therefore, accurate determination of solvent– and protein–ligand potentials is an important part of ligand entropy calculation.

In this report we present a fast method for estimating ligand entropy based on a Hessian matrix built during optimization with a quasi-Newton optimization. The basic assumption we made is that ligand molecules exist in different conformations in the gas or solution phases and that their fractional numbers in those phases are given by a Boltzmann distribution. This might not be a reasonable assumption for the ligand poses bound in the protein receptor; in this case, we consider a single binding mode given by the crystal structure of the protein–ligand complex.

In the next section, a description of our method is given, followed by its validation against gas- and solution-phase entropies. The last section contains a comparison of experimental and calculated values of $T\Delta S$ for the process of a selected ligand's binding by four protein receptors. We demonstrate that this method is comparable in precision to the more accurate determination of the vibrational part of entropy based on the exact Hessian and is suitable for use in rapid evaluations of ligand entropies.

Methods

Configurational entropy of a ligand in the gas- and solution-phases is calculated from[1]

$$S_c = kN\left[1 + \ln\left(\frac{q}{N}\right) + \frac{T}{q}\frac{\partial q}{\partial T}\right] \quad (3)$$

where k is the Boltzmann constant, N is the number of ligand molecules, T is the absolute temperature, and q is the partition function:

$$q = q_t \sum_{i=1}^{n_c} e^{-\varepsilon_i/kT} q_{iv} q_{ir} \quad (4)$$

where q_t is the translational partition function, the summation is over the number of ligand conformers, n_c, q_{iv}, q_{ir} are vibrational and rotational partition functions of conformer i, and ε_i is a sum of the internal energy and solvation free energy of conformation i.

In the case of protein-bound ligands, we assume that three translational and three rotational degrees of freedom of a

ligand are transformed into six degrees of vibrational motion of a trapped ligand, so eq 4 is reduced to

$$q = \begin{cases} q_v & \text{for single binding mode} \\ \sum_{i=1}^{n_p} \exp\left(-\frac{\varepsilon_i}{kT}\right)q_{iv} & \text{for multiple binding modes} \end{cases} \quad (5)$$

where n_p is the number of binding modes and q_{iv} is the vibrational partition function of a bound ligand in mode i.

Translational entropy in solution was calculated from the Sackur–Tetrode equation:

$$S_t = Nk\left(\ln\frac{1}{\rho\Lambda^3} + \frac{5}{2}\right) \quad (6)$$

where $\rho = N/V$ is the number density of the ligand in solution (set at 1 M concentration when molar entropy was evaluated) and Λ is the thermal de Broglie wavelength dependent on the mass of ligand molecule m and temperature:

$$\Lambda = \frac{h}{(2\pi mkT)^{1/2}} \quad (7)$$

No empirical correction to S_t was applied in order to account for the finite volume of solute molecules. The entropic effects of the excluded volume (cavity) not available for the solvent are an important part of solvation entropy and can be explicitly included by adding appropriate solvation terms. We have chosen to use this approach in our calculations (see Solvation Entropy Section). In order to use effectively the above formulation of ligand configurational entropy, we need three fast and reliable computational procedures:

(i) A method for generating an ensemble of ligand conformations.
(ii) A method for determining vibrational frequencies of each ligand conformer.
(iii) Method for estimating solvation effects on solute entropy.

Each method is described briefly below.

Conformation Generation. We have generated conformer ensembles by a method of random coordinates embedding, MMFF94 force field[13] refinement of fragments, combining of fragments into a molecule and finally torsion driving of rotatable bonds as implemented in Omega (version 2.1).[14,15] Conformations were generated using default parameters, except an 0.1 Å root-mean-square (RMS) threshold was used to determine and eliminate duplicate conformations. This low limit is intended to assume that the majority of conformers are included in entropy calculations. All conformations generated in this manner were energy minimized with the MMFF94 force field but only structurally unique conformations that differed after minimization by at least 0.05 Å in root-mean-square deviation (RMSD) are included in the entropy calculations. The importance of thorough conformation sampling will be addressed in Results and Discussion Section.

Vibrational Frequencies Determination. Our computational procedure introduces the following approximations: (1) that the vibrational motion of a molecule is represented

as a set of independent, uncoupled oscillators and (2) that each of those oscillators is harmonic. We realize that one might expect that the low-frequency motion of a ligand in the protein binding site might be significantly anharmonic and that the issue of associated error in entropy will need to be addressed in future, but far more important is the quality of the molecular potential that determines the shape of the potential well at minimum and, therefore, the values of the vibrational frequencies. For the purpose of this study, we adopted the MMFF94 potential force field,[13] which recently has been proved by us to have performed well in the refinement of crystallographic ligand structures when fitted into their electron densities.[16]

In order to calculate the vibrational entropy, we derived the vibrational frequencies from the normal-mode analysis of mass-weighted Hessian matrix of second derivatives:

$$\mathbf{H}^m = \mathbf{M}^{-1/2}\mathbf{H}\mathbf{M}^{-1/2} \tag{8}$$

where $\mathbf{H}$ and $\mathbf{M}$ are the Hessian and atomic mass matrices, respectively. Eigenvalues λ_i of $\mathbf{H}^m$ determine harmonic wavenumbers $\tilde{\nu}_i$ following eq 9:

$$\tilde{\nu}_i = \frac{1}{2\pi c}\sqrt{\lambda_i} \tag{9}$$

where c is the speed of light.

When the molecular potential is fully analytic, the Hessian matrix can be calculated for the optimized ligand structure. However, in order to make calculations more efficient and extend them for cases where some potential terms are given in the form of a high-resolution lookup table or three-dimensional (3D) grid, this work uses a Hessian matrix built by quasi-Newton type optimizers for the purpose of predicting the next step:

$$\mathbf{x}_{i+1} = \mathbf{x}_i + \mathbf{H}_{i+1}^{-1}\nabla f(\mathbf{x}_i) \tag{10}$$

where $\mathbf{x}_i$ is the coordinate vector at step i, $\mathbf{H}_{i+1}^{-1}$ is an inverse of the Hessian matrix at step $i + 1$ constructed according to the adopted scheme (for example, the widely used Broyden−Fletcher−Goldfarb−Shanno, BFGS, scheme),[17] and $f(\mathbf{x})$ is a function for which minimum is sought. At every iteration the approximate $\mathbf{H}^{-1}$ is closer to the exact Hessian, so when the optimization procedure is converged, a good quality Hessian can be obtained. Before the normal frequencies are calculated, we make sure that the Hessian is stable and does indeed determine a real minimum. If one or more eigenvalues are negative, the geometry of the ligand is randomly perturbed and reoptimized using the final matrix $\mathbf{H}^{-1}$ from the previous optimization as an initial guess of the invert Hessian. We have found that such a correction is needed very rarely when the molecular potential is fully analytic and, more frequently but still only occasionally, when some potential terms are evaluated on a grid.

The procedure outlined above is expected to result in frequencies and moments of inertia reasonably close to experiment only for the gas phase or for small ligands in solution with no rotatable bonds. The structure of highly flexible polar ligands, however, may be largely changed in solution. That implies a need to include solvent forces in

order to properly estimate vibrational and rotational entropy in solution. Here, a simple analytical solvation model recently developed by Grant and colleagues, known as the Sheffield solvation model,[18] is adopted. Briefly, the solvation energy of a ligand in solution is expressed as

$$E_s = -\frac{f_\varepsilon}{8\pi\varepsilon_0}\sum_{i,j}\frac{Q_iQ_j}{\sqrt{ar_ir_j + bR_{ij}^2}} \tag{11}$$

where ε_0 is permittivity of vacuum, Q_i and Q_j are partial charges on atoms i and j, r_i and r_j are atomic radii, R_{ij} is the distance between atoms i and j, and the factor f_ε is defined by dielectric constants of the solvent and the ligand (solute):

$$f_\varepsilon = \left(\frac{1}{\varepsilon_{solu}} - \frac{1}{\varepsilon_{solv}}\right) \tag{12}$$

Dimensionless parameters a and b have been chosen in such a way that the solvation energy given by eq 11 agrees with a physically rigorous model based on the Poisson equation.

Solvation Entropy. Solvation of a ligand in solution is associated with the formation of cavities around solute molecules and the reorganization of water molecules around them due to electrostatic and nonelectrostatic solute−solvent interactions. Each of those phenomena are accompanied by entropy changes. Comparison of estimated ligand entropy in solution with that of the corresponding experimental data, therefore, requires inclusion of solvation entropy ΔS_s (along with solute configurational entropy S_c) in the expression for the total ligand solution entropy:

$$S_{solution} = S_c + \Delta S_s \tag{13}$$

ΔS_s consists of electrostatic and hydrophobic parts:

$$\Delta S_s = \Delta S_{s,elec} + \Delta S_{s,hyd} \tag{14}$$

The way we estimate both parts of solvation entropy is given below.

Electrostatic Solvation Entropy. Bulk effects of entropy change upon solvation due to the electrostatic properties of the solution result from the temperature dependence of the solvent dielectric constant. This portion of solvation entropy is estimated from

$$\Delta S_{s,elec\_bulk} = -\left(\frac{\partial\Delta G_s}{\partial\varepsilon_{solv}}\right)\left(\frac{\partial\varepsilon_{solv}}{\partial T}\right) \tag{15}$$

where ε_{solv} is the solvent dielectric constant. The first term in eq 15 can be calculated from the simple analytical solvation model, while the second term is calculated from the experimental temperature dependence of the water dielectric constant.[19]

Entropic effects that reflect the change of the water structure upon solvation, in particular formation of hydrogen bonds between the solute and the solvent, are difficult to estimate and are beyond the methods that treat water as dielectric continuum. Water is a highly structured liquid in which every H_2O molecule with two proton donor and two proton acceptor sites forms four hydrogen bonds (H-bonds)

with its neighbors in a tetrahedral arrangement. Entropy change for the formation of such a single H-bond in water is estimated to be -37.2 J/(mol K).[20] Solute molecules containing H-bond donors or acceptors engage water molecules from the first solvation shell in solute−solvent H-bond formation, thus perturbing the original highly organized liquid structure. Also, the vibrational entropy of the H-bonded solute·$(H_2O)_n$ complexes is usually higher than that of the corresponding water clusters due to the larger number of low-frequency modes available in the former.

In order to account for the formation of ligand−solvent clusters due to strong hydrophilic interactions, we propose using of an empiric constant term of 28 J/(mol K) that increases the solution entropy of a compound containing proton donors or acceptors. The rationale of such a choice is based on the following two observations:

(i) Calculations of the vibrational entropy change for a number of gas-phase reactions:

$$(H_2O)_2 + M \rightarrow M \cdot H_2O + H_2O$$
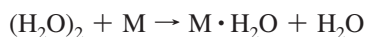
for a number of molecules M that contain proton donors or acceptors with the use of the MMFF94 potential yield a vibrational entropy change in the range 15−50 J/(mol K). One might expect, therefore, a small favorable entropic effect enhancing solubility of such compounds.

(ii) Our calculations of solution entropy for polar molecules or molecules with π-electron donors produce consistently underestimated values by 10−45 J/(mol K), with the remarkable exception of alcohols, for which a satisfactory agreement between calculated and experimental solution entropies is observed. One explanation for this observation is that in all cases but alcohols we are missing the entropy change of the H-bond rearrangements upon solvation in the first solvation shell. It is possible that the hydroxyl group in alcohols, with a very similar charge distribution to the hydroxyl group in water, does not significantly perturb the water structure even in the first solvation shell, while the increase of the vibrational entropy upon ROH·H_2O formation is small.

The proposed entropy correction accounting for solute−solvent clusters formation due to specific hydrophilic interactions is certainly a crude approximation, however, its accurate calculation for drug-like ligands is too difficult at this stage of our method development.

Hydrophobic Solvation Entropy. Calculation of the hydrophobic portion of solvation entropy is usually performed by one of a few approximate methods. We estimate this solvation entropy term in two ways. One is based on the common assumption that the free energy of hydrophobic solvation is proportional to the solute molecular surface area

$$\Delta G_{s,hyd} = \gamma A \tag{16}$$

where γ is the microscopic surface tension coefficient. The value of γ is in the range of 0−10 cal/(mol Å)[21,22] when $\Delta G_{s,hyd}$ represents the overall hydrophobic solvation effect, which includes both cavitation and van der Waals solute−

solvent interactions. We may, therefore, assume that γ is made of two components: $\gamma = \gamma_{cav} + \gamma_{vdw}$. Assuming that $\Delta G_{s,hyd}$ depends linearly on temperature and that its enthalpic contribution does not depend on temperature, the corresponding entropy change is

$$\Delta S_{s,hyd} = -\frac{\Delta G_{s,hyd}}{T} \tag{17}$$

Under the above assumptions, $\Delta G_{s,hyd}$ in eq 17 represents the temperature-dependent cavity formation term for which the value of $\gamma = \gamma_{cav}$ is about 30 cal/(mol Å), as determined for alkanes.[23,24] Using a set of 294 molecules containing a variety of functional groups, we have found that $\gamma = 30$ cal/(mol Å) indeed returns the best agreement on average with solution entropies.

The main criticism of using eq 16 is that for small molecules the hydrophobic portion of entropy is primarily related to the creation of a cavity in the solvent and scales with its volume[25] rather than the molecule's surface area A. In addition, coefficient γ has no precisely established value for all compounds. One has to conclude, therefore, that the widely used expression (eq 16), although useful, is not physically sound.

For the purpose of comparison, we apply, therefore, an alternative approach in which the hydrophobic free energy change $\Delta G_{s,hyd}$ is evaluated as a sum of a cavity formation component, van der Waals solute−solvent interaction and inductive (permanent dipole−induced dipole) interaction:

$$\Delta G_{s,hyd} = \Delta G_{cav} + \Delta G_{vdw} + \Delta G_{ind} \tag{18}$$

The first component in eq 18 is evaluated from the scaled particle theory (SPT)[9,26,27] according to the expression containing up to the cubic term:

$$\Delta G_{cav} = RT[K_0 + K_1(\sigma_c/\sigma_s) + K_2(\sigma_c/\sigma_s)^2 + K_3(\sigma_c/\sigma_s)^3] \tag{19}$$

where $K_0 = -\ln(1 - \xi)$, $K_1 = 3\xi/(1 - \xi) = u$, $K_2 = u(u + 2)/2$, $K_3 = \xi P v_s/RT$, σ_c and σ_s are cavity and solvent diameters, respectively (where solvent and solute are assumed spherical), ξ is the packing density of the solvent, and v_s is the solvent molar volume. The second term in eq 18 is calculated according to Pierotti's relationship:[26]

$$\Delta G_{vdw} = -(64/3)\xi\varepsilon_{ls}(\sigma_{ls}/\sigma_c)^3 \tag{20}$$

where ε_{ls} is Lennard-Jones (LJ) potential depth parameter for ligand−solvent interaction calculated as $(\varepsilon_l\varepsilon_s)^{1/2}$ where ε_l and ε_s are the corresponding parameters for ligand and solvent, respectively. The last term in eq 18, in most cases very small (of the order of 0.1 J/(mol K)) in comparison to ΔG_{cav} and ΔG_{vdw}, was evaluated according to[26]

$$\Delta G_{ind} = -8\xi(\sigma_c\sigma_{ls})^{-3}(\mu_l^2\alpha_s + \mu_s^2\alpha_l) \tag{21}$$

where μ_s, α_s, μ_l, and α_l are solvent and ligand dipole moment and polarizability, respectively. The value of σ_{ls} is taken as the average $(\sigma_s + \sigma_c)/2$. $\Delta S_{s,hyd}$ is determined from

$$\Delta S_{s,hyd} = \Delta S_{cav} - \left(\frac{\partial \Delta G_{vdw}}{\partial T} + \frac{\partial \Delta G_{ind}}{\partial T}\right) \quad (22)$$

where the entropy change for cavity formation ΔS_{cav} is calculated from ΔG_{cav} (eq 19) and the corresponding enthalpy change

$$\Delta H_{cav} = [\xi\alpha RT^2/(1-\xi)^3][(1-\xi)^2 + 3(1-\xi)(\sigma_c/\sigma_s) + $$
$$3(1+2\xi)(\sigma_c/\sigma_s)^2] + \xi P v_s(\sigma_c/\sigma_s)^3 \quad (23)$$

as

$$\Delta S_c = (\Delta H_{cav} - \Delta G_{cav})/T \quad (24)$$

Temperature derivatives of ΔG_{vdw} and ΔG_{ind} are determined from the known experimental temperature dependence of σ_s, ξ, and v_s for water. Quantity α is the thermal expansion coefficient of the solvent.

The accuracy of all the thermodynamic quantities derived from the SPT theory (eqs 19–24) depends critically on solvent parameters. Values adopted in this work for water are: molecular volume $v_s = 18.0685$ cm^{3}mol^{-1}, thermal expansion coefficient $\alpha = 2.572 \times 10^{-4}$ K^{-1}, effective hard sphere diameter $\sigma_s = 2.8$ Å, LJ potential depth $\varepsilon_s/k = 100$ K, dipole moment $\mu_s = 1.8$ D, and polarizability $\alpha_s = 1.4573$ Å^3. The value of the hard sphere diameter for water of 2.8 Å results from the location of the first peak in the oxygen−oxygen radial distribution function for water,[28] while the LJ parameter of 100 K (ratio of Lennard-Jones parameter ε_s to Planck's constant, to be correct) was adopted after Graziano[9] as the value that works well for alkanes and alcohols.

In the case of solute parameters, we applied the simple procedures described below to estimate all necessary parameters (σ_c, ε_l/k, μ_l, and α_l). For molecules with multiple conformations, we used the geometry of the most stable conformation in solution optimized with the MMFF94 and Sheffield forces.

Hard Sphere Diameters. For each solute, the hard sphere parameter was calculated from its molecular Gaussian volume V_G[29] as $2(0.75V_G/\pi)^{1/3}$. We found a fair agreement between hard sphere diameters calculated in this way with the values obtained from solubility data for a number of molecules listed in Table 1.

LJ Potential Depth. Experimental LJ potential parameters derived from either second virial coefficients or viscosity measurements are available only for a small number of molecules. Critically evaluated data based on the second type of measurements are available for 75 simple molecules in the 1977 paper of Mourits and Rummens.[31] Two decades later, Cuadros et al.[32] developed a simulation-based procedure for evaluating LJ parameters for any small compound, which was recently used by Cachadina and Mulero for calculating the vaporization enthalpies for over 1500 substances.[33] Our procedure of assigning the LJ potential depth parameter of a compound uses both above-mentioned sources of data in a hierarchical fashion: if the compound happens to belong to the Mourits and Rummens set, the corresponding experimental value is used; otherwise the data obtained by Cachadina and Mulero are adopted. If a compound is not found in any of the above two sets, an average value of 370 K for σ_c/k is used. This represents the average of the Cachadina and Mulero set excluding mono-, di-, and triaomic species.

Dipole Moments. The dipole moment for the most stable solution conformation calculated from its AM1BCC partial charges was used. Because such a value usually overestimates the experimentally measured dipole moment, we further scale it by a factor of 0.82, which is the ratio of experimental-to-calculated dipole moment for water molecule. Other molecules are known to have similarly overpolarized charges with the AM1BCC charging method.

Molecular Polarizabilities. The molecular polarizability of a compound was estimated according to the recently published method of Wang et al.[34] (model 2E in Table 4 in that publication).

Results and Discussion

When conformation ensembles are used for the estimation of entropy, an important issue is the completeness of those ensembles. The lack of thorough comformation sampling can be illustrated with the gas-phase *n*-nonane: using an RMS threshold of 0.8 Å for duplicate removal results in 25 conformations, which, after minimization, appear to be unique. The total calculated entropy of *n*-nonane at 298 K for this set of conformations is 452.9 J/(mol K). The use of a 0.1 Å RMS threshold produces 130 conformations which after minimization are reduced to 128 structurally unique conformers and results in 472.7 J/(mol K) total entropy. The error of about 20 J/(mol K) (or 4%) can easily be eliminated by a fine-grain sampling of the conformational space. Further lowering of the duplicate removal threshold to an RMS of 0.005 Å has no effect: 239 generated conformations contained only 128 structurally unique conformers after optimizations.

Gas-Phase Molecules. Gas-phase entropies calculated with the MMFF94 force field and quasi-Newton Hessian frequencies for the vibrational entropy component for most small molecules with no torsions or containing only an isolated methyl group bonded to aromatic rings (like toluene) are in very good agreement with experimental values. This is shown in Figure 1a for a number of molecules containing carbon, oxygen, sulfur, nitrogen, and halogen atoms. In contrast, molecules with single torsional bonds show up to 8% error in calculated entropies with respect to experimental values. This behavior is shown in Figure 1b for *n*-alkanes from ethane to decane and is not surprising given the underlying approximations described in the previous section. In particular, we might expect that the torsional, low-frequency vibrations are significantly anharmonic, so the error in entropy estimation for larger, more flexible molecules will be larger than for rigid ones. A good illustration of poorer entropy estimation for flexible and satisfactory predictions for rigid molecules are calculated values for benzene and *n*-hexane of 270.2 and 367.3 J/(mol K), respectively, and the corresponding experimental values of 269.2 and 388.4 J/(mol K). Similarly, the RMS deviation between calculated and experimental values in the series of

Table 1. Comparison of Hard Sphere Diameters Derived from Gaussian Molecular Volumes (σ_G)[a]

compound	σ_G	σ_{solu}	compound	σ_G	σ_{solu}
carbon dioxide	4.10	3.94	benzene	5.29	5.26
methane	3.60	3.70	toluene	5.61	5.64
ethane	4.20	4.38	*m*-xylene	5.89	5.97
ethylene	4.07	4.07	fluorobenzene	5.34	5.30
n-hexane	5.73	5.92	chlorobenzene	5.71	5.61
n-heptane	6.01	6.25	hexafluorobenzene	5.59	5.65
n-octane	6.26	6.54	nitrobenzene	5.85	5.74
n-nonane	6.50	6.83	methanol	4.01	3.69
n-decane	6.71	7.08	ethanol	4.53	4.34
3-methylheptane	6.26	6.52	cyclohexanol	5.76	5.75
2,3-dimethylhexane	6.26	6.50	acetone	4.86	4.76
cyclohexane	5.57	5.63	*N*-methylacetamide	5.15	4.96
methylcyclohexane	5.86	5.99	dimethylsulfoxide	5.10	4.91

[a] Corresponding values obtained from solubility data by Wilhelm and Battino[30] (σ_{solu}). All values in Å.
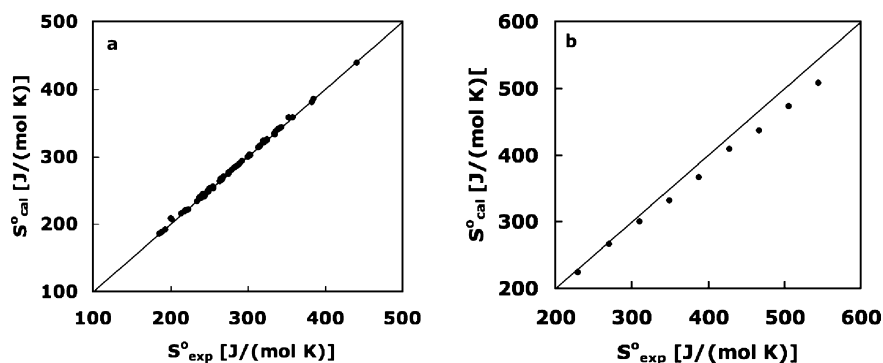

Figure 1. Calculated versus experimental entropies at 298K for the below-listed gas-phase molecules. Plot a: methane,[a] ethylene,[a] acetylene,[a] cyclopropane,[b] allene,[b] benzene,[b] toluene,[b] naphtalene,[b] water,[a] carbon dioxide,[a] formaldehyde,[a] formic acid,[a] oxirane,[a] methanol,[b] phenol,[b] thiacyclopropane,[b] thiophene,[b] methanethiol,[b] ammonia,[a] hydrogen cyanide,[a] aziridine,[b] pyridine,[b] nitric acid,[a] nitrous acid,[a] acetonitrile,[b] benzoninitrile,[b] nitromethane,[b] methyl nitrate,[b] 3-picoline,[b] fluoromethane,[a] tetrafluoro-ethylene,[b] 1,1-difluoroethylene,[b] trifluoroethylene,[b] fluorobenzene,[b] hexafluorobenzene,[b] *p*-fluorotoluene,[b] chloromethane,[a] chloroethylene,[b] tetrachloroethylene,[a] 1,1-dichloroethylene,[b] *t*-1,2-dichloroethylene,[b] trichloroethylene,[b] chlorobenzene,[b] 1,2-dichlorobenzene,[b] 1,3-dichlorobenzene,[b] 1,4-dichlorobenzene,[b] hexachlorobenzene,[b] bromomethane,[a] bromoethylene,[b] bro-momethane,[b] chlorotrifluoroethylene,[b] iodomethane,[b] iodobenzene.[b] Plot b: *n*-alkanes from ethane to *n*-decane. [a]Experimental values are taken from ref 35, and [b]experimental values are taken from ref 36. Solid diagonal lines show ideal behavior of calculated values.

thiacycloalkanes from thiacyclopropane to thiacycloheptane is 5.7 J/(mol K), while the corresponding value for 1-thiols from ethanethiol to 1-hexanethiol is 22.4 J/(mol K).

Anharmonicity error in the calculation of normal frequencies has been recognized for a long time in quantum chemical calculations; a number of remedies, including scaling frequencies by factors within the (0.8, 1.0) range,[37] selective or overall scaling of force constants,[38] and estimation of anharmonicity constants by numerical calculation of third and fourth potential derivatives along the normal modes,[39,40] have been proposed and successfully used to predict the gas-phase IR and Raman frequencies of a variety of small molecules. For the purpose of our method, we adopted a simple frequency scaling. Based on a set of 255 molecules for which standard gas-phase entropies are available in the compilation of Domalski and Hearing,[36] we obtained an optimum scaling factor of 0.85. Figure 2, compares calculated gas-phase standard entropies with scaled and unscaled frequencies with the experimental (or recommended) values for the above-mentioned set of 255 molecules. This figure shows that scaling of frequencies results in much better agreement with the published experimental or recommended entropies. Unfortunately, for some molecules the scaling

Table 2. Comparison of Molecular Diameters Obtained in Two Ways[a,b]

compound	experimental	unscaled	scaled
cyclohexanone	322.2	341.1	354.5
thiacyclopentane	361.9	371.3	389.9
hexachlorobenzene	441.2	439.1	463.8

[a] σ_G - hard sphere diameters derived from Gaussian molecular volume, and σ_{solu} - diameters obtained from the solubility data by Wilhelm and Battino.[30] [b] All values in A.

procedure results in significant overestimation of entropy that is particularly visible in the range of 300−500 J/(mol K). For example, visibly deviated points from the ideal line correspond to cyclohexanone, thiacycloheptane, and hexachlo-robenzene, for which the experimental and calculated standard entropies in J/(mol K) with unscaled and scaled normal frequencies are see in Table 2.

It is seen that for these compounds, frequency scaling increases the error by largely overestimating the calculated entropy.

Another source of error is the approximate character of the quasi-Newton Hessian matrix. As mentioned in the Vibrational Frequencies Determination Section, a good
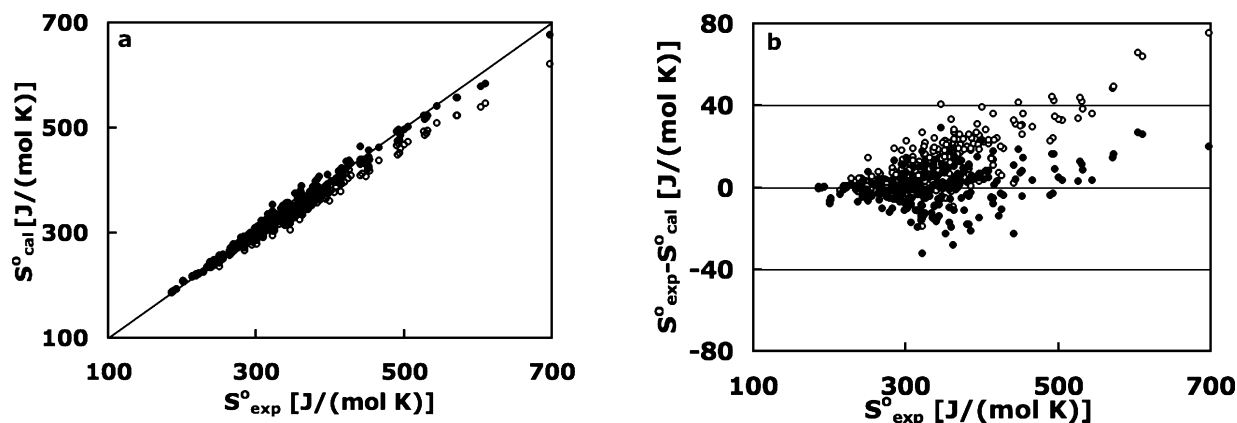
Figure 2. Gas-phase standard entropies (plot a) and their displacements from experimental values (plot b) calculated with the use of unscaled (○) and scaled (●) normal frequencies for a set of 255 molecules containing carbon, oxygen, sulfur, nitrogen, and halogen atoms. A list of molecules other than those mentioned in the caption of Figure 1 is given in the Appendix Section. The RMS displacement from published entropies is 18.2 and 9.4 J/(mol K) for unscaled and scaled normal freqiencies, respectively.
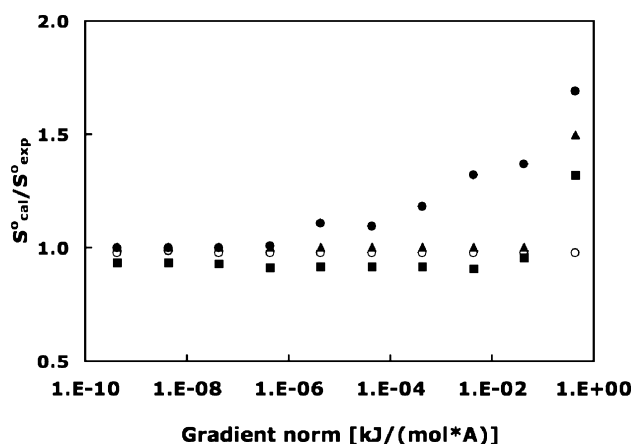

Figure 3. Ratio of calculated-to-experimental gas-phase entropy for ethane (○), *n*-decane (□), benzene (▲), and cyclohexane (●) as a function of gradient norm convergence criteria in quasi-Newton built Hessian.

quality approximation of the quasi-Newton Hessian matrix is obtained only when the optimization process is converged. Figure 3 shows the effect of convergence criteria on the calculated gas-phase entropy of four hydrocarbon molecules. It is seen that, although for a molecule with no torsions, like benzene, gradient norm reduction to 10^{-3} kJ/(mol Å^2) is sufficient to obtain a stable converged value of entropy, the data for cyclohexane suggest that a convergence criteria of at least 10^{-7} kJ/(mol Å^2) is necessary to minimize the error related to the approximated Hessian obtained in the quasi-Newton optimization. Therefore, in all entropy calculations we use a convergence on gradient norm below that value, typically 10^{-10} kJ/(mol Å^2).

In order to estimate the effect of the approximate character of the quasi-Newton converged Hessian, we repeated the calculations of entropies for the same set of 255 compounds from Figure 2 but with the exact, analytical Hessian matrix calculated for the optimized compounds. The RMS displacement of entropies calculated in such a way from published values is 15.0 J/(mol K), compared to 20.8 J/(mol K) obtained with the quasi-Newton frequencies. The improvement due to the diagonalization of the exact Hessian is,

therefore, real (as might be expected) but on average is not dramatic. The quality of the force field is probably of more significance, but in this paper, we make no attempt to evaluate different available molecular potentials for entropy estimation.

Molecules in Solution. The goal of this study is to elucidate a very rapid means of calculating ligand entropies in gas and condensed phases. Thus, we sought an approximate solvation model that could rapidly evaluate solvent forces that modify equilibrium geometries and normal frequencies of ligand conformations in solution. To determine the accuracy of ligand entropies calculated with this approximate solvent model, we compared the values calculated with a more rigorous but significantly slower Poisson model. Calculated configurational entropies for 20 drug molecules in solution with the use of a simplified analytical solvation Sheffield model[18] in comparison to a Poisson model are shown in Figure 4a. It is seen that for the majority of molecules, the Sheffield values are slightly larger. The average signed error of Sheffield entropies is 13.9 J/(mol K). Observed differences might also, however, be impacted by the larger numerical errors in the calculation of the quasi-Newton Hessian in the case of the Poisson potential evaluated on the grid with respect to the fully analytical Sheffield potential. Figure 4b compares the CPU times of both types of entropy calculations and shows the obvious speed advantage in the case of Sheffield model. Large CPU times visible on the in Figure 4b for the Poisson-type entropy calculations reflect numerous reoptimizations (as outlined in the Methods Section) needed for some conformations to meet the formal requirements for a Hessian matrix. Differences in configurational entropy calculated with the two solvent models are small (3−6%), yet on average, the Sheffield model is over two orders of magnitude faster to calculate. We conclude that using the Sheffield solvation model for rapid entropy estimation of solution molecules is a reasonable and beneficial approach.

Total solution entropies calculated according to eq 13 can be compared with the corresponding experimental entropy
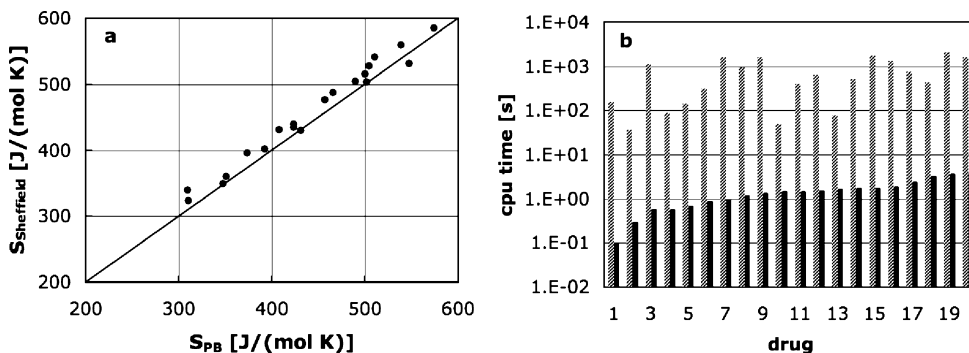
Figure 4. (a) Calculated configurational entropies for 20 drug molecules using Sheffield solvation model ($S_{Sheffield}$) vs Poisson model (S_{PB}). (b) CPU times used for entropy calculations for each drug using Sheffield model (solid bars) and Poisson model (shaded bars). Numbers on horizontal axis correspond to the following drugs: varenicline (1), acetaminophen (2), aspirin (3), ephedrine (4), telbiduvine (5), lamivudine (6), ofloxacin (7), decitabine (8), meloxicam (9), hydrocortizone (10), sertraline (11), ciproflaxin (12), desonide (13), ibuprofen (14), zoledronic acid (15), neralabine (16), ritalin (17), venlaxafine (18), warfarin (19), and tramadol (20).
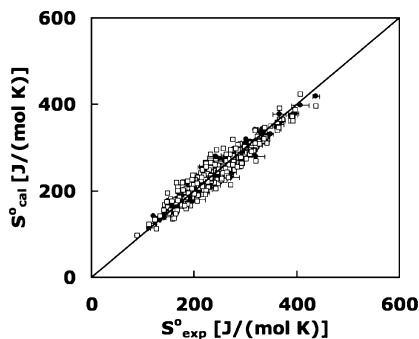

Figure 5. Calculated vs experimental standard molar entropies for 294 molecules in solution representing the following group of compounds: normal and branched alkanes, cycloalkanes, alkenes, alkynes, alkylbenzenes, alcohols, ketones, ethers, esters, thiols, sulfides, nitriles, and amines. Experimental data are taken from the ORCHYD database[41] except those for amines and three purines (adenine, xanthine, and hypoxanthine). Experimental data for amines are taken from the 1981 paper of Cabani et al.[42] Data for purines are taken from papers of Boyer et al.[43] and Tewari et al.[44] Full circles (●) show calculated standard entropies with the use of the SPT theory, while squares (□) show similar data calculated with the SA eqs 16 and 17 with tension coefficient γ set at 30 cal/(mol/ Å). The values of R^2 and slopes (R^2,slope) are (0.945,0.923) and (0.878,0.865), respectively. Corresponding lower and upper 95% confidence intervals for R^2, obtained from Fisher transformation are (0.931,0.956) for SPT and (0.849,0.902) for SA solvation models, respectively. Straight line shows ideal agreement with experiment.

Figure 6. Calculated vs experimental standard molar entropies for 244 molecules, a subset of molecules for which data are presented in Figure 5. Calculations were done with the SPT/Sheffield (●) and SPT/Poissson (△). The corresonding values of R^2 and slopes are (0.911,0.906) and (0.899,0.808), respectively. The 95% confidence intervals for R^2 are (0.887,0.930) and (0.872, 0.921), respectively.

data. Figure 5 shows such a comparison for 294 molecules containing a variety of functional groups.

It is seen in Figure 5 that the calculated values follow the experimental values; however, the use of the SPT theory returns a better agreement with experimental data, as is visible from slopes, R^2's, and confidence intervals for R^2. The root-mean-square error (RMSE) of the calculated standard solution entropies using the SPT model from the corresponding experimental data is 13.6 J/(mol K), while the largest deviation visible on the plot is the underestimation of the molar solution entropy by 39.8 J/(mol K) for 1,8-nonadiyne. The corresponding values for the SA model are 20.0 and 56.0 J/(mol K) overestimation for trans-1,4-dimethylcyclohexane. The differences in prediction accuracy
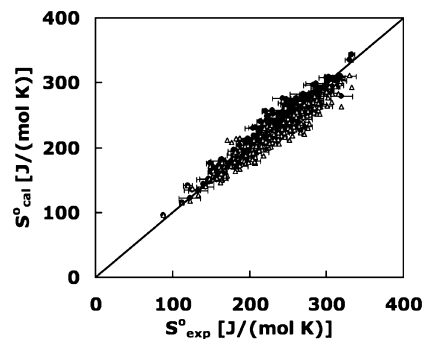
between the two models are not dramatic but statistically significant: larger R^2 for the SPT method (0.95 vs 0.88) together with the lack of overlap of 95% confidence intervals for R^2, shown in the caption for Figure 5, clearly illustrates that the SPT model outperforms the SA method in prediction of solution entropies. We should also mention that our SPT method has no adjustable parameters, whereas the SA method was optimized albeit to a value that could have been estimated from the physical properties of alkane−water transfer.

Figure 6 shows the comparison between total solution entropies calculated with Sheffield/SPT and physically more rigorous PB/SPT model. There is little difference in the use of the Sheffield or the full PB model; while the slopes are slightly different, the correlation coefficients are identical within statistical error of 95% confidence. This is a remarkable result, indicating that a simple, fully analytical, and fast solvation model generates the same quality entropy prediction for molecules in solution that the physically correct but much more expensive to use Poisson solvation model. As discussed above, slightly better prediction of Sheffield/SPT method is not surprising and can be attributed to the higher accuracy of analytical second-order derivatives.

We hope that using a molecular potential of higher quality than MMFF and refinement of the model describing solvent−solute interactions contribution to the solvation

terms (both vdw and hydrogen bonding) (eq 14) could lead to a better agreement with experimental data. It was shown that, with water as a solvent, SPT provides reliable results for simple small molecules,[45,46] however, we are not aware of its prior application to complex drug-like compounds. An extended version of SPT for solutes of arbitrary shapes has been developed,[47] but it is not clear if it generates significantly better results in aqueous solutions than its original hard-sphere version, so at this point, we do not have plans to implement it just for the purpose of entropy calculations.

Protein-Bound Molecules. The only way to test our method of estimated ligand entropy inside a protein receptor is to compare calculated and experimental entropy change ($T\Delta S$) upon ligand binding. Only limited amounts of such experimental data from microcalorimetry experiments exist. In addition, cases where there is a significant contribution to the overall entropy change from the protein itself, caused by large change in protein structure upon binding, cannot be used for the purpose of testing because we are not attempting here to estimate protein entropy change. In fact, our basic approximation in evaluating a protein-bound ligand is the complete rigidity of a protein. Such a crude approximation, therefore, severely restricts the scope of protein–ligand complexes which might be handled by our method.

We have chosen four protein–ligands systems shown below. In selecting these systems, our main criteria were lack of favorable entropy change upon binding ($T\Delta S > 0$) (protein contribution to $T\Delta S$ cannot be ignored otherwise), moderate size of ligands, quality of experimental data, and of course, existence of the protein–ligand X-ray structures in the Protein Date Bank (PDB):[48]

(i) Major urinary protein (MUP-I)–*n*-alcohols (1znd, 1zne, 1zng, 1znh, and 1znk).
(ii) Renin–diaminopyrimidines (2iko, 2iku, and 2il2).
(iii) Aldose reductase–sorbinil/fidarestat (2pdk and 1pwm).
(iv) Quinone reductase–resveratrol/melatonin (1sgo and 2qx4).

X-ray protein structures used for calculations are shown above. Protein preparation included hydrogenation to standard residues ionization states and optimization of hydrogen atom positions with the MMFF94 force field in vacuum. Entropy change upon ligand binding has to include partial solvation of that part of the ligand in the protein–ligand complex which is exposed to the solvent, $f\Delta S_s$, and partial desolvation of the protein binding site, ΔS_{des}, upon ligand binding.

$$\Delta S_{bind} = S_{protein} - S_{solution} + f\Delta S_s + \Delta S_{des} \quad (25)$$

ΔS_s is given by eq 14, and the fraction f of a bound ligand surface exposed to the solvent is evaluated from the molecular surfaces area of a ligand A_L, protein A_P, and protein–ligand complex A_{PL}:

$$f = 0.5(A_L - A_P + A_{PL})/A_L \quad (26)$$

Entropy of partial protein desolvation, ΔS_{des}, is assumed to be caused by the change of the hydrophobic part of protein desolvation and is calculated from the surface area expression (eq 17) because for macromolecules it scales with the molecular surface.[25] The microscopic surface tension coeffiecient γ is set at 5 cal/(mol Å).

MUP-I–*n*-Alcohols. Structures for the complexes of pheromone-binding protein MUP-1 with five primary alco-

Table 3. Calculated and Experimental Values of $-T\Delta S$ for the Binding Equilibria MUP-I + *n*-Alcohol $\rightleftharpoons$ MUP-I·*n*-Alcohol[a]

alcohol	$-T\Delta S_{cal}$	$-T\Delta S_{exp}$
pentanol	18.4, 22.6[b]	17.9 ± 3.2
hexanol	32.3, 21.0[b]	19.3 ± 0.6
heptanol	30.6, 19.9[b]	20.9 ± 0.4
octanol	25.7	22.4 ± 0.6
nonanol	26.1	24.8 ± 0.5

[a] $-T\Delta S$ in kJ/mol. [b] Value for the alternative position. [c] Value for double binding modes.

hols: pentanol, hexanol, heptanol, octanol, and nonanol were taken from the PDB entries 1znd, 1zne, 1zng, 1znh, and 1znh, respectively. The protein structures in these models are essentially unchanged in each complex, but ligands can be bound in two structurally different positions depicted on Structure I. In the case of the pentanol complex, Malham et al. have observed both positions simultaneously, however, the second one (mode II) with much weaker density.[49] In the first run of entropy calculations, we assumed a single binding mode with starting geometries given in the above crystal structures. Table 3 contains the results along with the experimental results of isothermal titration microcalorimetry (ITC) performed by Malham et al.[49] It is seen that the calculated entropy penalty agrees very well with the experimental values for pentanol and fairly well with the corresponding values for octanol and nonanol. For hexanol and heptanol, our prediction is too negative by 13.0 and 9.7 kJ/mol, respectively. In order to explain the difference, we repeated the calculations for those two alcohols and for pentanol, assuming that the initial position for pentanol is given by the alternative position observed experimentally by Malham et al.,[49] while the starting poses for hexanol and heptanol were obtained by truncating the terminal carbons from the octanol 1znh structure (mode II in structure I). Recalculated $-T\Delta S$ values for pentanol, hexanol, and heptanol are marked in Table 3 with a footnote (b, in this case). It is seen that those latter values for both hexanol and heptanol are in very good agreement with the experimental data of Malham et al.[49] In the case of pentanol, the alternative pose leads to larger entropy penalty and larger deviation from the experimental value (-22.6 vs -18.4 kJ/mol).
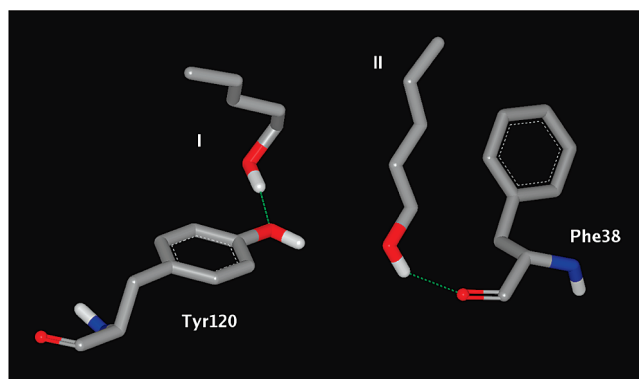
Structure I

Our calculations suggest that as alcohol molecules grow, due to steric hindrance only, mode II seems to be available,

Ligand Entropy in Gas-Phase

J. Chem. Theory Comput., Vol. 6, No. 7, 2010 **2149**

Table 4. Calculated and Experimental Values of $-T\Delta S$ for Several Binding Equilibria Protein + Ligand ⇌ Protein·Ligand[a]

| protein	compound	$-T\Delta S_{cal}$	$-T\Delta S_{exp}$
renin	I	12.6	8.4 ± 4.2
	II	7.5	-1.0 ± 4.2
	III	3.3	1.8 ± 4.2
aldose	IV	11.7	13.9 ± 1.6
reductase	V	13.1	28.8 ± 0.8
	IV[b]	19.7	
	V[b]	19.9	
quinone	VI	23.1	21.0 ± 2.6
reductase 2	VII	10.0	10.0 ± 0.5

[a] Reported experimental error for renin binding are from a personal communication of R.W. Sarver. $-T\Delta S$ in kJ/mol. [b] Anion bound by a protein with protonated His110.

Compound I

Compound II

Compound III

while for smaller pentanol, both are likely to occur with the majority binding in mode I. X-ray data for pentanol, octanol, and nonanol complexes are in full agreement with the above suggestion. The source of disagreement for hexanol and heptanol is not clear. One possibility is that our model does not include explicit water molecules in the binding site. Such water molecules were observed in X-ray data by Malham et al.,[49] and because they form H-bonds with hydroxyl groups of both bound alcohols and Tyr120, they may contribute to binding entropy. On the other hand, distribution between two different modes of binding in the crystal and the liquid phase might be different, so our prediction based on entropy calculation should be verified with nuclear magnetic resonance (NMR), rather X-ray structures. Finally there is a possibility that a higher resolution X-ray structure determination of the protein−ligand complexes discussed here will show a complete agreement with our entropy calculations.

Renin−Diaminopyrimidines. Calculations were done for three ligands (compounds I−III), assuming single binding modes determined by the ligand coordinates in the corresponding PDB structures 2iko, 2iku, and 2il2.

Data in Table 4 show that, given the experimental uncertainty of 4 kJ/mol, predicted values of $-T\Delta S$ for compounds I and II are in fair agreement with the experimental results of Sarver et al.[50] For compound II, experiment shows the lack of entropy penalty (actually a small favorable effect) upon binding. Our calculations showing the entropy penalty of 7.5 kJ/mol does not reproduce this finding. It is likely that the loss of ligand entropy when the ligand binds to renin is compensated for the simultaneous increase in protein entropy, and as mentioned above, our current model does not include such effects.

Aldose Reductase−Sorbinil/Fidarestat. The only difference between sorbinil (compound IV) and the drug fidarestat (compound V) is the amide group missing in the former. Its oxygen forms a H-bond with the backbone NH group of Leu300 in the human enzyme. Fidarestat is bound tighter than sorbinil, which makes the former an efficient inhibitor of aldose reductase, able to slow the progression of diabetic neuropathy.[51] Microcalorimetry results of Petrova et al.[52] reported in Table 4 suggest however that the entropy penalty upon binding is twice as big as in the case of sorbinil. Our calculations based on the assumption of single binding modes for both compounds given in the crystal structures 1pwm (for fidarestat) and 2pdk (for sorbinil) do not confirm that finding. Predicted values of binding entropy $-T\Delta S$ are
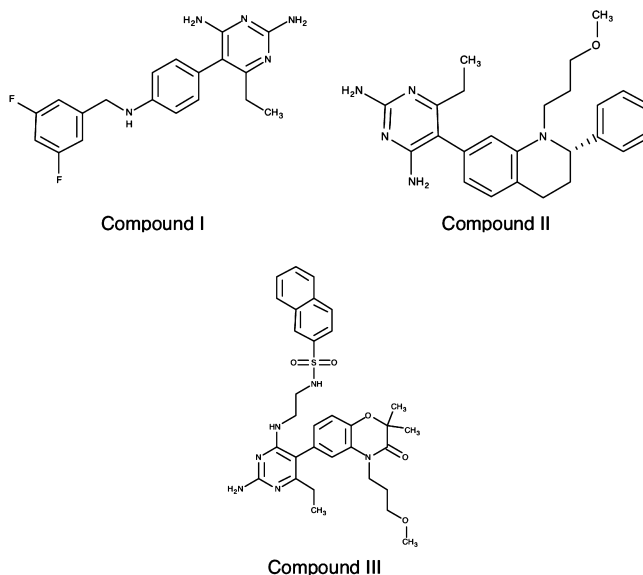
comparable for both compounds and agree well with the experimental value for sorbinil. More recent, very high-resolution (0.78 Å) crystallography data of Zhao et al.[53] offer a likely explanation of the discrepancy. It has been found that the bound ligand is deprotonated at the imide nitrogen and that the adjacent catalytic His110 is protonated. In other words, a strong electrostatic cation−anion attraction not only contributes to the large protein−ligand interaction (measured ΔH of binding is -75.5 kJ/mol[52]) but also decreases the entropy of both the ligand and the side chain of His110 by restricting their motion. Two possible mechanisms could lead to the formation of the observed salt bridge: first, Zhao et al.[53] suggest a two step process in which the binding of a neutral ligand is followed by a proton-transfer reaction, and second, a direct binding of an anionic form of fidarestat by the enzyme with a protonated His110. The second mechanism cannot be ruled out because the estimated fidarestat pK_a is in the range of 7.9−8.5 pH units (estimation was made using the data for phenytoin and analogs),[54] which indicates that at the pH = 8, at which the ITC experiments were done,[52] the fraction of ionized fidarestat in solution is at least 30%. We used our entropy calculation method to evaluate the binding entropy of the anionic form of fidarestat according to the above two scenarios. First, we calculated the entropy of a deprotonated fidarestat in the active site of aldose reductase with positively charged His110. The result of this calculation enabled us to estimate the entropy change for a proton-transfer reaction:

$$\text{aldose\_reductase}(\text{His}_{110}) \cdot \text{HL} \rightarrow$$
$$\text{aldose\_reductase}(\text{His}_{110}\text{H}^+) \cdot \text{L}^-$$
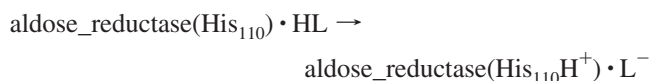
where HL is a neutral form of fidarestat. The estimated value of $T\Delta S$ at room temperature for the above reaction is -6.7 kJ/mol, so the overall calculated entropy change upon binding is $(-13.1 - 6.7) = -19.8$ kJ/mol, much closer to the experimental value of -28.8 kJ/mol. Second, the entropy change for direct binding of anionic forms of both fidarestat and sorbinil were calculated, and the results shown in Table 4. It is seen that the calculated values for the anionic form of fidarestat are essentially identical for both mechanisms,

so our method does not discriminate between the two possibilities. For both mechanisms of fidarestat anion binding, the calculated values are significantly more negative than in the case of the neutral ligand. Still more negative by about 9 kJ/mol, the experimental value probably reflects the restriction of protonated His110 motion in the electrostatic field of the negatively charged ligand, but as we mentioned earlier, our current model is not able to capture any protein entropy change contribution. The result for sorbinil suggests that the neutral form of the latter is bound by aldose reductase.

Compound IV   Compound V
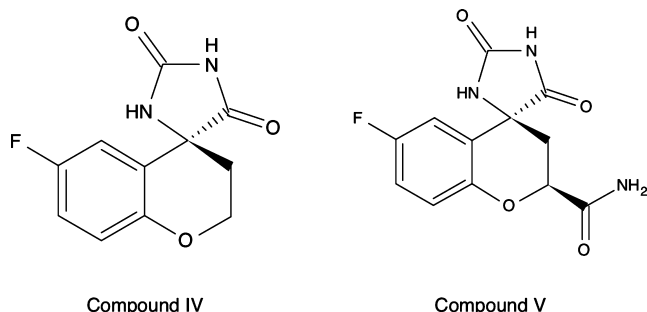
Quinone Reductase 2—Resveratrol/Melatonin. Structures of resveratrol, a natural polyphenol found in wine, and neurohormone melatonin are depicted as compounds VI and VII. The molecules bind in such a way that their aromatic rings are positioned in parallel to the isoalloxazine ring of a cofactor FAD. Initial positions of both ligands and the corresponding QR2 protein coordinates were taken from the PDB entries 1sg0 and 2qx4, respectively. Calculated and experimental values derived from ITC experiments of Calamini et al.[55] are shown in Table 4. Calculated values of $-T\Delta S$ for both resveratrol and melatonin are in very good agreement with experiment. Although given the approximate model used in our calculations, the ideal agreement for melatonin seems to be fortuitous.

Compound VI   Compound VII
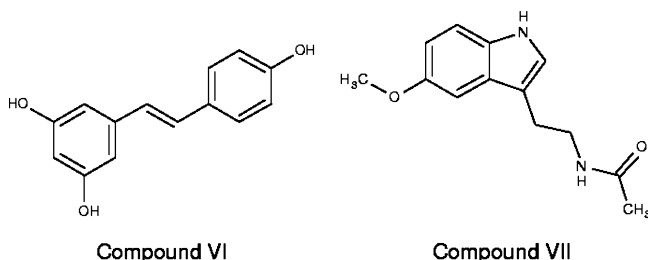
Binding Entropy, Summary. Figure 7 shows the calculated binding entropies vs experimental ITC data for all tested protein—ligands complexes. The squared correlation coefficient ($R^2 = 0.81$) is relatively high, however, its confidence limits are large. In addition, considering that the calculated correlation is based only on those binding modes for hexanol and heptanol, which produce the closest agreement with experiment and preselection of those protein—ligand complexes for which protein contribution to the binding entropy is presumably negligible or low, at this moment, we are not able to evaluate the reliability of the method for binding entropy prediction. Instead we are considering these preliminary results as encouraging for further improvement of a high-throughput method for binding entropy prediction. Data for aldose reductase/fidarestat show the importance of both ligand and protein charge states,
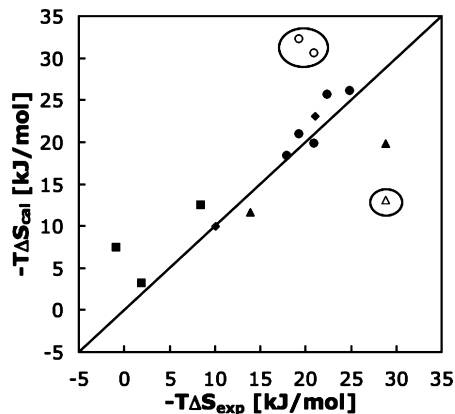
Figure 7. Calculated vs experimental binding entropies for protein—ligand systems selected for current study: (MUP-I)—*n*-alcohols (●), renin—diaminopyrimidines (■), aldose reductase—sorbinil/fidarestat (▲), and quinone reductose—resveratrol/melatonin (◆). Straight line represent ideal agreement. $R^2 = 0.81$, and its 95% confidence interval is (0.45–0.94). Points in ovals represent calculated data for hexanol and heptanol in binding mode I (structure I) and neutral form of fidarestat, respectively, and are not included to correlation.
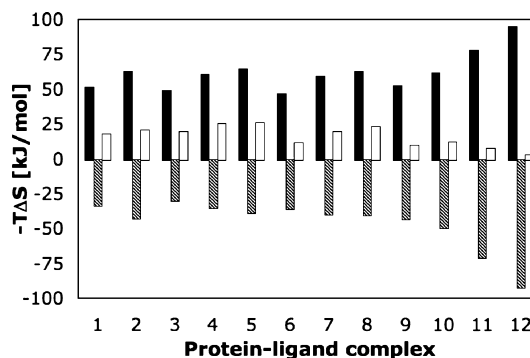
Figure 8. Calculated configurational entropy change (solid bars) and total solvation entropy change (shaded bars) upon binding. Empty bars show the $-T\Delta S_{bind}$ values from Figure 7. Numbers on horizontal axis correspond to protein—ligand complexes: MUP-I—pentanol (1), MUP-I—hexanol (2), MUP-I—heptanol (3), MUP-I—octanol (4), MUP-I —nonanol (5), aldose reductase—sorbinil (6), aldose reductase—fidarestat (7), quinone reductase—resveratrol (8), quinone reductase—melatonin (9), renin—compound I (10), renin—compound II (11), and renin—compound III (12).

while data for MUP-I/hexanol and MUP-I/heptanol show the necessity of considering multiple binding modes.

Given the correlation shown in Figure 7, it is tempting to search for a dominant calculated entropy term standing behind it. Figure 8 shows the configurational and the total solvation entropy change upon binding for the same 12 protein—ligand complexes which produced the correlation in Figure 7.

It is seen that both configurational and solvational entropy components play equally important roles in the binding entropy, so neither configurational entropy nor solvation entropy change alone is responsible for the correlation shown in Figure 7. In some protein—ligand complexes (like in the case of MUP-I complexed with alcohols), the loss of the configurational entropy is significantly larger than the entropy gain resulting from the solvation entropy change, while in

other cases (renin complexed with diaminopyrimidines), both components largely compensate for each other which results in little or no entropy penalty upon binding.

The ultimate goal of this work is to provide an accurate estimate of the entropy of binding of a small, drug-like molecule to a protein. Such an estimate is either time-consuming and difficult to calculate, for instance requiring the extensive sampling available from molecular dynamics, or is provided by very crude methods, such as penalty terms from the number of rotatable bonds. What we attempt here is to explore whether methods of intermediate speed and complexity can none the less be accurate. If so, it will provide a significant component of the long sought for computational approach to affinity prediction and its application to drug discovery.

Conclusions

- The Hessian matrix generated by the quasi-Newton optimizers of molecular structure with the use of a good quality force field can be used to predict vibrational entropy in the gas-phase, solution, and protein receptor environments.
- The simplified analytical solvation model formulated recently by Grant et al.[18] can be used for ligand entropy calculations in solution.
- Entropy of molecules containing rotatable bonds is underestimated probably due to anharmonicity of the low-frequency torsional movements.
- The use of scaled particle theory (SPT) for the determination of the total solution entropies of compounds is promising. It eliminates the uncertainty associated with the value of surface tension coeffiecient used in the surface area (SA) model of hydrophobic solvation.
- Further work is required to address the anharmonicity of the quasi-Newton frequencies, the still unsatisfactory hydrophobic solvation model, and the contribution of protein receptors structural changes to the ligand binding entropy changes. This includes the usage of more accurate potential (force field, atomic charges on protein receptor, and analysis of protein ionization states), and entropic effects due to hydrogen bond and nonelectrostatic (van der Waals) solute−solvent interactions. Estimation of protein entropy with the use of a quasi-Newton Hessian matrix will require the usage of more efficient optimization techniques, such as limited memory Broyden−Fletcher−Goldfarb−Shanno (BFGS) technique.
- Our results for aldose reductase−fidarestat system strongly suggest that pK_a analysis of both ligand and protein receptor should be included in the entropy estimation for protein−ligand binding.

Appendix

List of molecules which in addition to those listed in the caption for Figure 1 were used to determine the frequency scaling factor in entropy calculations.

Hydrocarbons.

Ethane, propane, *n*-butane, *n*-pentane, *n*-hexane, *n*-heptane, *n*-octane, *n*-nonane, *n*-decane, 2-methylpropane, 2-methylbutane, 2-methylpentane, 2-methylhexane, 3-methylhexane, propylene, 2,2-dimethylpropane, 2,3-dimethylbutane, 2,2-dimethylbutane, 1-butene, 1-pentene, *trans*-2-butene, *cis*-2-butene, *trans*-2-pentene, *cis*-2-pentene, 1,2-butadiene, 1,3-butadiene, propyne, 1-butyne, 2-butyne, butadiyne, biphenyl, 1,2-dimethylbenzene, 1,2,3-trimethylbenzene, 1,2,4-trimethylbenzene, pentamethylbenzene, and hexamethylbenzene.

Oxygen Compounds.

Ethanol, 2-propanol, *n*-propanol, *n*-butanol, *tert*-butyl alcohol, cyclohexanol, allyl alcohol, 2-butanol, *o*-cresol, *m*-cresol, *p*-cresol, dimethyl ether, diethyl ether, di-*n*-propyl ether, di-*n*-butyl ether, methyl−ethyl ether, methyl−propyl ether, tetrahydrofuran, 1,4-dioxan, acetaldehyde, propanal, acetone, butanal, pentanal, methyl−ethyl ketone, methyl−propyl ketone, diethyl ketone, cyclohexanone, acetic acid, methyl formate, and ethyl acetate.

Sulfur Compounds.

Ethanethiol, 1-propanethiol, 1-butanethiol, 1-pentanethiol, 1-hexanethiol, 1-heptanethiol, 1-octanethiol, 1-nonanethiol, 1-decanethiol, 2-propanethiol, 2-butanethiol, cyclopentanethiol, 2-methyl-1-propanethiol, 2-methyl-2-propanethiol, 2-methyl-2-butanethiol, benzenethiol, diethyl sulfide, dimethyl sulfide, ethyl−methyl sulfide, isopropyl−methyl sulfide, propyl−methyl sulfide, butyl−methyl sulfide, propyl−ethyl sulfide, butyl−ethyl sulfide, diisopropyl sulfide, pentyl−methyl sulfide, dipropyl sulfide, butyl−propyl sulfide, pentyl−ethyl sulfide, hexyl−methyl sulfide, dibutyl-sulfide, hexyl−ethyl sulfide, heptyl−methyl sulfide, dipentyl sulfide, *t*-butyl−methyl sulfide, diethyl-disulfide, dimethyl-disulfide, dipropyl-disulfide, dibutyl-disulfide, dimethyl-sulfoxide, dimethyl-sulfone, thiacyclobutane, thiacyclopentane, thiacyclohexane, thiacycloheptane, 2-methylthio-phene, and 3-methylthiophene.

Nitrogen Compounds.

Methylamine, ethylamine, *n*-propylamine, *n*-butylamine, *n*-pentylamine, *n*-hexylamine, ethylenediamine, 2-aminobutan, *t*-butylamine, dimethylamine, diethylamine, trimethylamine, triethylamine, aniline, propionitrile, butyronitrile, acrylonitrile, hydrazine, pyrrolidine, 2-picoline, nitroethane, nitropropane, nitrobutane, 2-nitropropane, 2-nitrobutane, ethyl nitrate, *n*-propyl-nitrate, and isopropyl-nitrate.

Halogen Compounds.

Fluoroethane, 1-fluoropropane, 1,1-difluoroethane, 1,1,1-trifluoroethane, hexafluoroethane, 1-chloropropane, chloroethane, 1-chlorobutane, 1-chloropentane, 2-chloropropane, 2-chlorobutane, 1-chloro-3-methylbutane, 1-chloro-2-methylpropane, 2-chloro-2-methylpropane, 1,2-dichloropropane, 1,2-dichloroethane, 2-chloro-2-methylbutane, 1,3-dichloropropane, 1,1-dichloroethane,

1,1,2-trichloroethane, 2,2-dichloropropane, 1,2,3-trichloropropane, 1,1,2,2-tetrachloroethane, 3-chloro-1-propene, pentachloroethane, hexachloroethane, acetylchloride, bromoethane, 1-bromopropane, 1-bromobutane, 1-bromopentane, 2-bromopropane, 2-bromobutane, 2-bromo-2-methylpropane, 1,2-dibromoethane, 1,2-dibromopropane, 1,2-dibromobutane, 2,3-dibromobutane, 2,3-dibromo-2-methylbutane, 3-bromo-1-propene, 1-bromopropyne, iodoethane, 1-iodopropane, 2-iodo-2-methylpropane, 2-iodopropane, 1,2-diiodoethane, 1,2-diiodopropane, 1,2-diiodobutane, and 3-iodo-1-propene.

References

(1) McQuarie, D. *Statistical Mechanics*; Harper & Row: New York, 1976; pp 129−149.

(2) Edholm, O.; Berendsen, H. *Mol. Phys.* **1984**, *51*, 1011–1028.

(3) Karplus, M.; Kushick, J. *Macromolecules* **1981**, *14*, 325–332.

(4) Andricioaei, I.; Karplus, M. *J. Chem. Phys.* **2001**, *115*, 6289–6292.

(5) Schlitter, J. *Chem. Phys. Leters* **2003**, *215*, 617–621.

(6) Mammen, M.; Shakhnovitch, E.; Deutch, J.; Whitesides, G. *J. Org. Chem.* **1998**, *63*, 3821–3830.

(7) Amzel, L. *Proteins* **1997**, *28*, 144–149.

(8) Siebert, X.; Amzel, L. *Proteins* **2004**, *54*, 104–115.

(9) Graziano, G. *J. Phys. Chem. B* **2005**, *109*, 12160–12166.

(10) Chang, C.; Gilson, M. *J. Am. Chem. Soc.* **2004**, *126*, 13156–13164.

(11) Chen, W.; Chang, C.; Gilson, M. *Biophys. J.* **2004**, *87*, 3035–3049.

(12) Chang, C.; Chen, W.; Gilson, M. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 1534–1539.

(13) Halgren, A. *J. Comput. Chem.* **1996**, *17*, 490–519, 520−552, 553−586, 587−615, 616−641.

(14) *Omega2.1*, OpenEye Scientific Software, Inc.: Santa Fe, NM, 2006.

(15) Boström, J. *J. Comput.-Aided Mol. Des.* **2001**, *15*, 1137–1152.

(16) Wlodek, S.; Skillman, A.; Nicholls, A. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2006**, *62*, 741–749.

(17) Nocedal, J.; Wright, S. *Numerical Optimization*; Glynn, P., Robinson, S. M., Eds.; Springer-Verlag: New York, 1999.

(18) Grant, J.; Pickup, B.; Sykes, M.; Kitchen, C.; Nicholls, A. *Chem. Phys. Lett.* **2007**, *441*, 163–166.

(19) *CRC Handbook of Chemistry and Physics*; Lide, D., Ed.; CRC Press: New York, 1999.

(20) Suresh, S.; Naik, V. *J. Chem. Phys.* **2000**, *113*, 9727–9732.

(21) Reynolds, J.; Gilbert, D.; Tanford, C. *Proc. Natl. Acad. Sci. U.S.A.* **1974**, *71*, 2925–2927.

(22) Sitkoff, D.; Sharp, K.; Honig, B. *Biophys. Chem* **1994**, *51*, 397–409.

(23) Hermann, R. *Proc. Natl. Acad. Sci. U.S.A.* **1977**, *74*, 4144–4145.

(24) Sharp, K.; Nicholls, A.; Fine, R.; Honig, B. *Science* **1991**, *252*, 106–109.

(25) Chandler, D. *Nature* **2005**, *437*, 640–647.

(26) Pierotti, R. *Chem. Rev.* **1976**, *76*, 717–726.

(27) Graziano, G. *J. Chem. Soc., Faraday Trans.* **1998**, *94*, 3345–3352.

(28) Head-Gordon, T.; Hura, G. *Chem. Rev.* **2002**, *102*, 2651–2670.

(29) Grant, J.; Gallardo, J.; Pickup, B. *J. Comput. Chem.* **1996**, *17*, 1653–1666.

(30) Wilhelm, E.; Batino, R. *J. Chem. Phys.* **1971**, *55*, 4012–4017.

(31) Mouritz, F.; Rummens, F. *Can. J. Chem.* **1977**, *55*, 3007–3020.

(32) Cuadros, F.; Mulero, A.; Cachadina, I. *Int. Rev. Phys. Chem.* **1995**, *14*, 205–213.

(33) Cachadina, I.; Mulero, A. *J. Phys. Chem. Ref. Data* **2007**, *36*, 1133–1139.

(34) Wang, J.; Xie, X.; Hou, T.; Xu, X. *J. Phys. Chem. A* **2007**, *111*, 4443–4448.

(35) *NIST Chemistry WebBook, NIST Standard Reference Database Number 69*; Linstrom, P.; Mallard, W., Eds.; National Institute of Standards and Technology: Gaithersburg, MD; http://webbook.nist.gov. Accessed January 30, 2010).

(36) Domalski, E.; Hearing, E. *J. Phys. Chem. Ref. Data* **1993**, *22*, 805–1159.

(37) Blom, C.; Altona, C. *Mol. Phys.* **1976**, *31*, 1377–1391.

(38) Rauhut, G.; Pulay, P. *J. Phys. Chem.* **1995**, *99*, 3093–3100.

(39) Barone, V. *J. Chem. Phys.* **2004**, *120*, 3059–3065.

(40) Barone, V. *J. Chem. Phys.* **2005**, *122*, 014108-1−014108-10.

(41) Plyasunova, N.; Plyasunov, A.; Shock, E. *Int. J. Thermophys.* **2004**, *25*, 351–360.

(42) Cabani, S.; Gianni, P.; Mollica, V.; Lepori, L. *J. Solution Chem.* **1981**, *10*, 563–595.

(43) Boyer, J.; Francis, M.; Boerio-Goates, J. *J. Chem. Thermodyn.* **2003**, *35*, 1917–1928.

(44) Tewari, Y.; Gery, P.; Vaudin, M.; Mighell, A.; Klein, R.; Goldberg, R. *J. Chem. Thermodyn.* **2004**, *36*, 645–658.

(45) Lee, B. *Biopolymers* **1991**, *31*, 993–1008.

(46) Graziano, G. *Biophys. Chem.* **2003**, *104*, 393–405.

(47) Irisa, M.; Nagayama, K.; Hirata, F. *Chem. Phys. Lett.* **1993**, *207*, 430–435.

(48) Berman, H.; Henrick, K.; Nakamura, H. *Nat. Struct. Biol.* **2003**, *10*, 980–980.

(49) Malham, R.; Johnstone, S.; Bingham, R.; Barratt, E.; Phillips, S.; Laughton, C.; Homans, S. *J. Am. Chem. Soc.* **2005**, *127*, 17061–17067.

(50) Sarver, R. *J. Anal. Biochem.* **2007**, *360*, 30–40.

(51) Hotta, N.; Toyota, T.; Matsuoka, K.; Shigeta, Y.; Kikkawa, R.; Kaneko, T.; Takahashi, A. *Diabetes Care* **2001**, *24*, 1776–1782.

(52) Petrova, T.; Steuber, H.; Hazemann, I.; Cousido-Siah, A.; Mitschler, A.; R.Chung,; Oka, M.; Klebe, G.; El-Kabbani, O.; Joachimiak, A.; Podjarny, A. *J. Med. Chem.* **2005**, *48*, 5659–5665.

(53) Zhao, H.; Hazemann, I.; Mitschler, A.; Carbone, V.; Joachimiak, A.; Ginell, S.; Podjarny, A.; El-Kabbani, O. *J. Med. Chem.* **2008**, *51*, 1478–1481.

(54) Porter, R.; Meldrum, B. *Basic and Clinical Pharmacology*, 5th ed.; Appleton & Lange: Norwalk, CT, 1992.

(55) Calamini, B.; Santarsiero, B.; Boutin, J.; Mesecar, A. *Biochem. J.* **2008**, *413*, 81–91.