Loss of Translational Entropy in Molecular Associations

ARTICLE *in* PROTEINS STRUCTURE FUNCTION AND BIOINFORMATICS · NOVEMBER 2003

Impact Factor: 2.63 · DOI: 10.1002/prot.10472 · Source: PubMed

2 AUTHORS, INCLUDING:

Xavier Siebert
Université de Mons
39 PUBLICATIONS **191** CITATIONS

SEE PROFILE

Loss of Translational Entropy in Molecular Associations.

Xavier Siebert and L. Mario Amzel†

Department of Biophysics and Biophysical Chemistry,
Johns Hopkins University School of Medicine,
Baltimore, MD 21205, USA

Molecular associations in solution are opposed by the loss of entropy (ΔS) that results from the restriction of motion of each of the components in the complex. Theoretical estimates of ΔS are essential for rationalizing binding affinities as well as for calculating entropic contribution to enzyme catalysis.

Recently a statistical-mechanical framework has been proposed for estimating efficiently the translational entropy loss (ΔS^{trsl}), while taking explicitely into account the complex intermolecular interactions between the solute and the solvent. This framework relates the translational entropy of a solute in solution to its "free volume", defined as the volume accessible to the center of mass of the solute in the presence of the solvent and calculated using an extension of the cell model (CM) for condensed phases. The translational entropy of pure water, estimated with the CM algorithm, shows good agreement with the experimental information. The free volume of various solutes in water, calculated within the CM, using molecular dynamics simulations with explicit solvent, displays a strong correlation with the solutes' polar and total surface areas. This correlation is used to propose a parameterization that can be used to calculate routinely the translational entropy of a solute in water.

We also applied the CM formalism to calculate the free volume and translational entropy loss (ΔS^{trsl}) upon binding of benzene to a cavity in a mutant T4-lysozyme. Our results agree with previously published estimates of the binding of benzene to this mutant T4-lysozyme. These and other considerations suggest that the cell model is a simple yet efficient theoretical framework to evaluate the translational entropy loss upon molecular association in solution.

I. INTRODUCTION

A Formulation of the problem

Molecular association in solution is opposed by the loss of entropy (ΔS) that results from the restriction of motion (translational, rotational, vibrational, conformational) of each of the components upon formation of the complex. Evaluation of ΔS from first principles is crucial for understanding ligand binding and for dissecting entropic contributions to enzyme catalysis [1–10].

The noncovalent association of two molecules A and B *in solution* (e.g., A = ligand, B = protein) to form a complex AB in the reaction $A + B \leftrightharpoons AB$ results in an entropy loss:

$$\Delta S = S_{AB} - S_A - S_B \tag{1}$$

The notation S_ξ ($\xi = A$, B or AB) represents the entropy of the species ξ *in solution* and depends on the concentration. ΔS is generally calculated at a standard (fictitious) concentration of $1M$ to allow comparison between different systems on a common basis [1, 11–13].

†correspondence to : mario@neruda.med.jhmi.edu

Using classical statistical mechanics, we can express the *translational* entropy S_ξ^{trsl} of a given concentration of solute ξ in solution (i.e., N_ξ solute and N_w solvent molecules in a volume V at a temperature T) from the molecular partition function of *one solute*, q_ξ, defined as [14]

$$q_\xi = \frac{1}{h^3} \int\int e^{-\beta H(\mathbf{p}_\xi, \mathbf{q}_\xi, \{\mathbf{q_w}\})} d\mathbf{p}_\xi d\mathbf{q}_\xi \tag{2}$$

where h is Planck's constant and β=1/kT. The integral is performed on the whole phase space of the solute, with $\mathbf{q}_\xi$ the position of its center of mass and $\mathbf{p}_\xi$ its momentum. The notation $\{\mathbf{q_w}\}$ represents the positions of all solvent molecules that create a potential field $U(\mathbf{q}_\xi, \{\mathbf{q_w}\})$ around the solute. Assuming the solution is dilute enough to neglect the solute-solute interactions, we obtain the Hamiltonian for the solute of mass m_ξ : $H(\mathbf{p}_\xi, \mathbf{q}_\xi, \{\mathbf{q_w}\}) = \frac{\mathbf{p}_\xi^2}{2m_\xi} + U(\mathbf{q}_\xi, \{\mathbf{q_w}\})$.

The translational entropy of the system of N_ξ solute molecules (S_ξ^{trsl}) is linked to the *total* partition function Q_ξ of the N_ξ solute molecules [14]:

$$S_\xi^{trsl} = k(\ln Q_\xi + T \frac{\delta \ln Q_\xi}{\delta T}|_V) \tag{3}$$

where k is Boltzmann's constant and $Q_\xi = q_\xi^{N_\xi}/N_\xi!$ if the N_ξ solutes are indistinguishable (e.g. gas or liquid state) or $Q_\xi = q_\xi^{N_\xi}/N_\xi^{N_\xi}$ otherwise (e.g. crystalline solid state or

bound molecule). Therefore, even at the same density, the entropy of N_ξ liquid molecules is higher than that of a solid containing the same number of molecules by kN_ξ. This term kN_ξ is sometimes called the "communal entropy" [15] and reflects the ability of molecules in a liquid to occupy any position in a given volume, as opposed to these in a solid whose positions are fixed [14].

Thus, the estimation of ΔS^{trsl} depends solely on the evaluation of the integral in Eq. 2 which, after integration over the momenta $d\mathbf{p}$, yields

$$q_\xi = \frac{1}{\Lambda_\xi^3} \int_V e^{-\beta U(\mathbf{q}_\xi, \{\mathbf{q_w}\})} d\mathbf{q}_\xi \tag{4}$$

where $\Lambda_\xi = \frac{h}{\sqrt{2\pi m_\xi kT}}$ is the De Broglie wavelength of a solute molecule [14].

If the N_ξ solutes were an *ideal gas* (i.e., without solvent nor intermolecular interactions, $U(\mathbf{q}_\xi, \{\mathbf{q_w}\}) = 0$), the integral in Eq. 4 would be equal to V, leading to the so-called "Sackur-Tetrode" (ST) formula for the translational entropy [14]:

$$S^{trsl}_{\xi,ST} = k(N_\xi \ln \frac{V}{N_\xi \Lambda_\xi^3} + \frac{5}{2} N_\xi) \tag{5}$$

This formula is sometimes used as the basis for estimating the entropies of molecular association in solution [3, 13, 16]. However, in solution the presence of a nonzero potential $U(\mathbf{q}_\xi, \{\mathbf{q_w}\})$ complicates considerably the evaluation of the configurational integral in Eq. 4, requiring approximate methods [1].

The goal of this paper is to provide a framework based on cell models [2, 15, 17, 18], to approximate efficiently ΔS^{trsl}.

B Cell Model for S^{trsl}_ξ

The cell model (CM) was originally proposed by Eyring, Hirschfelder and coworkers [15, 17] and developed more formally by Kirkwood [18]. Its extension to the problem of molecular association was recently suggested [2]. As the intermolecular interactions in solution are dominated by the short-range repulsive part of the potential $U(\mathbf{q}_\xi, \{\mathbf{q_w}\})$, the exponential factor in Eq. 4 vanishes except on a small portion of the space, which makes the calculation tractable.

The CM considers that the center of mass of each solute molecule is confined in a cell delimited by the field of the surrounding solvent molecules, $U(\mathbf{q}_\xi, \{\mathbf{q_w}\})$. In Eq. 4 we can divide the integration space into $N_\xi + N_w$ parts, and define the volume $v_{f,\xi}$ of a cell, sometimes called "free volume", as

$$v_{f,\xi} = \int_{\frac{V}{N_\xi + N_w}} e^{-\beta U(\mathbf{q}_\xi, \{\mathbf{q_w}\})} d\mathbf{q}_\xi \tag{6}$$

Eq. 4 leads to $q_\xi = (N_\xi + N_w) v_{f,\xi} / \Lambda_\xi^3$ and the translational entropy of the N_ξ solutes immersed in N_w solvent molecules becomes:

$$S^{trsl}_{\xi,CM} = k(N_\xi \ln \frac{(N_w + N_\xi) v_{f,\xi}}{N_\xi \Lambda_\xi^3} + \frac{5}{2} N_\xi) \tag{7}$$

In the limiting case of an ideal gas of N_ξ solutes without solvent ($N_w = 0$), the free volume is $v_{f,\xi} = V/N_\xi$ and we recover Eq. 5. The problem of evaluating the configurational integral in Eq. 4 is now reduced to the evaluation of the free volume using Eq. 6.

The main approximations inherent to the CM are twofold: 1) the division of the space in cells, each containing only one molecule and 2) the calculation of the free volume.

In the first part of this paper we calculate v_f in Eq. 6 using molecular dynamics (MD) simulations to evaluate directly the potential $U(\mathbf{q}_\xi, \{\mathbf{q_w}\})$ for a variety of solutes in water and discuss the results in terms of the polarity and size of each solute. We also use the CM to calculate the translational entropy for the particular case of liquid water (i.e., ξ is itself a water molecule), which we can compare with experimental information and with the Sackur-Tetrode (ST) estimate (Eq. 5).

C Cell Model for ΔS^{trsl}

The loss of translational entropy in Eq. 1 ($\Delta S^{trsl} = S^{trsl}_{AB} - S^{trsl}_A - S^{trsl}_B$) can be evaluated using an extension of the CM for S^{trsl}_ξ, described in the previous section. The term S^{trsl}_A and S^{trsl}_B are calculated directly using Eq. 7, with $\xi = A$ or $\xi = B$, respectively.

The term S^{trsl}_{AB} is more complicated, in that A and B are confined into a complex, which we will emphasize by denoting them A_c and B_c, respectively. We cannot simply calculate S^{trsl}_{AB} using Eq. 7 with $\xi = AB$, because this would correspond to considering only the motion of the (noncovalent) complex AB as a whole, neglecting the residual translational motion of B relative to A.

It is possible to use Eq. 7 to calculate $S^{trsl}_{A_c}$ and $S^{trsl}_{B_c}$, but the integral giving the free volume of B_c (in Eq. 6) must be restricted to the portion of space where B_c is bound to A_c. This also obliterates the "communal entropy" of B_c, lowering its entropy by a factor kN_{B_c}.

The CM formalism gives then finally the loss of translational entropy loss upon molecular association, expressed at a standard state of 1M of solute in 55M of water (N_{av} being Avogadro's number):

$$\Delta S^{trsl}_{CM} = kN_{av}(-\ln 56 + \ln \frac{v_{f,A_c} v_{f,B_c}}{v_{f,A} v_{f,B}}) - kN_{av} \tag{8}$$

Conservation of mass imposes cancelation of the terms corresponding to the integration over the momenta [1, 19].

As pointed out in [2], the last term of Eq. 8, $kN_{av} = R = 1.98\, e.u. (e.u. = \frac{cal}{mol.K})$ is sometimes used as an approximation of the translational entropy loss, assuming implicitly that the ratio of the free volumes in Eq. 8 is unity. This assumption will be discussed in the light of the results of this paper.

In this study we apply the CM formalism to calculate the free volume and translational entropy of a benzene molecule binding to a mutant T4-lysozyme (Eq. 8). We use a normal-mode analysis to demonstrate that the motion of benzene within the binding site of lysozyme can indeed be treated as a translation and compare our results with a recent estimate based on a body-restrain algorithm for MD simulation [20].

II. METHODS

A Molecular dynamics simulation

Estimation of the free volume of a solute ξ in a solvent, given by the integral in Eq. 6, requires an evaluation of the energy of interaction of the solute with the solvent, $U(\mathbf{q}_\xi, \{\mathbf{q_w}\})$. This energy must in principle be evaluated for all positions of the solute ($\mathbf{q}_\xi$) in the volume $\frac{V}{N_\xi + N_w}$ around its equilibrium position and averaged over a set of configurations of the surrounding solvent molecules $\{\mathbf{q_w}\}$ representative of the NVT ensemble.

Using molecular dynamics (MD) simulations to generate trajectories of the solute and the solvent molecules, we used individual "snapshots" to calculate the potential $U(\mathbf{q}_\xi, \{\mathbf{q_w}\})$. In all the simulations described below we used the CHARMm28 force field [21] and the SHAKE algorithm [22] to allow an integration time step of 2 fs, saving snapshots of configurations every 2 ps for analysis. Unless otherwise stated, we maintained the temperature at 300K by adjusting the kinetic energy term [23] and used a cutoff of 12 Å for non-bonded interactions.

1 Solute in water

Trajectories of pure water were generated first. 512 water molecules (TIP3 model) were placed in a cubic box of side length = 24.8346 Å, to get the density of liquid water at 300 K and 1 bar ($\rho \approx 1 g/cm^3$, equivalent to 55 M). As water molecules were treated explicitely, we used a dielectric constant of 1 for the electrostatic interactions. Cubic periodic boundary conditions were used, with a cutoff of 12 Å for the interactions with the periodic images. We minimized the total energy using 400 steps of steepest descent, heated the water box from 0 to 300K for 20 ps, thermalized it for 40 ps at 300K and generated 800 ps of trajectories at T=300K. We tested two different methods for maintaining the temperature at 300K to probe the sensitivity of the free volume calculations: adjusting the kinetic energy [23] and using a Nose-Hoover thermostat [24, 25].

We then inserted the solute (see Table I for the list of solutes used in this study) into the center of the box of water and into its periodic images, removing the water molecules that overlapped with the solute. As the 12 Å cutoff is slightly less than half the box side length, solute molecules do not interact with their images. This mimics an ideal, infinitely dilute solution [1, 11–13]. The standard state concentration of $1M$ solute is obtained by adjusting N_ξ and N_w in Eq. 7. We again minimized the total energy using 400 steps of steepest descent, first with the solute constrained to relax only the surrounding water, and then for 400 steps without restraints. We heated the system from 0 to 300K for 20 ps, thermalized all the molecules in the box for 40 ps at 300K, and produced 800 ps trajectories at 300K for each solute.

2 Benzene binding to a mutant T4-lysozyme

A mutant T4-lysozyme (L99A, C54T, C97A, hereafter called TCM, for Triple Cavity Mutant) was designed to create a hydrophobic cavity large enough to accommodate a benzene molecule [26]. The 1.9 Å crystal structure provided coordinates for 162 out of 164 residues of the TCM T4-lysozyme in complex with benzene (PDB entry: 1L84, [26]). Two missing amino acids, located at the C terminus, far from the cavity, are unlikely to affect the results. As the cavity is quite rigid and excludes solvent [20], these simulations were performed in vacuum. The structure was minimized for 400 cycles of steepest descent, and then subjected to a 60 ps equilibration step at 300K. Small harmonic constraints (2.4 Kcal/mol Å^2) force constant on the C_α atoms were imposed to avoid unfolding of the protein but still allow the ligand to move in the cavity during 800 ps of production at 300K.

B Free volume calculations

The integral in Eq. 6 was approximated as a summation over a three-dimensional fine grid (spacing $\delta_x = \delta_y = \delta_z = 0.1$Å) set up over the volume of the cell. The x, y and z axes of the grid coincide with the principal axes of inertia of the solute, so that all snapshots could be aligned. The potential $U(x, y, z, \{\mathbf{q_w}\})$ of interaction between the solute located at each grid point $\mathbf{q}_\xi = (x, y, z)$ and the surrounding water molecules, at positions $\{\mathbf{q_w}\}$ comprises the two nonbonded, distance-dependent terms : Van der Waals and electrostatics. It was calculated using as reference value the minimum within the cell, i.e. $\Delta U(x, y, z, \{\mathbf{q_w}\}) = U(x, y, z, \{\mathbf{q_w}\}) - U(\min_x, \min_y, \min_z, \{\mathbf{q_w}\})$. For each selected snapshot (i) we calculated

$$v^i_{f,\xi} \approx \sum_{x,y,z=-g}^{g} \exp[-\beta \Delta U^i(x, y, z, \{\mathbf{q_w}\})] \delta_x \delta_y \delta_z \qquad (9)$$

C Normal Mode Analysis

The motion of the bound ligand was also analyzed using the normal modes analysis facility ("vibran") of the CHARMm28 program [21] starting with the minimized structure of the

complex. The TCM T4-lysozyme was kept fixed and the normal modes were calculated in the reduced basis corresponding to the motion of the benzene in the field created by the fixed lysozyme. The six modes of lowest frequency corresponded, as expected, to translational/rotational motions. The energy profiles along these modes were explored and compared with those of the four *internal* modes with lowest frequency.

III. RESULTS AND DISCUSSION

A Solute ξ in water

The various solutes used in this analysis were chosen for their different sizes and patterns of surface polarities (see Table I).

The free volume of a solute ξ in water reflects restriction of the motion of its center of mass. This depends on the interaction energy $U(q_\xi, \{\mathbf{q_w}\})$ between the solute and the field created by the surrounding water molecules. To use Eq. 9, this interaction energy was calculated for each snapshot on a three-dimensional grid centered around the equilibrium position of the solute. Fig. 1a and Fig. 1c present a one-dimensional slice (along the x axis, in the y=z=0 grid plane) of the interaction energy $U(x, 0, 0, \{\mathbf{q_w}\})$ as well as a quadratic fit to this energy profile. This emphasizes the anharmonic character of the energy profile, for solutes as different as glycerol (Fig. 1a) and benzene (Fig. 1c). Fig. 1b and Fig. 1d show the 2D-profile of the exponential factor $\exp[-\frac{\Delta U(x,y,0,\{\mathbf{q_w}\})}{kT}]$ used in Eq. 9 and illustrate that the translation of the solute molecule in the field created by the surrounding water molecules is restricted to less than 1 Å in each direction, before the Boltzmann factor falls to zero. Due to its higher surface polarity, the motion of glycerol in water is more restricted than that of the nonpolar benzene.

The corresponding energy profiles for methylacetate (data not shown) are of the same order of magnitude and present intermediate features between that of glycerol and benzene, consistent with the respective surface polarities of these molecules.

The free volume v_f^i was computed for each snapshot i; a normalized probability distribution of v_f^i is plotted in Fig. 2 for three selected solutes, of comparable size (glycerol, methylacetate, benzene).

The distribution for glycerol (Fig. 2) shows a sharp peak at low v_f^i, reflecting the tight packing of water molecules hydrogen-bonded to the glycerol and a small tail at large v_f^i that extends to about 0.1 Å^3. In contrast, the histogram for benzene is broader, with a much smoother cutoff at low v_f^i and a wider tail at large v_f^i. The average free volume for benzene shifts towards higher values (see Table I). This reflects the looser packing of the clathrate-forming water molecules around the hydrophobic benzene molecule. For methylacetate, the situation is intermediate between the glycerol and

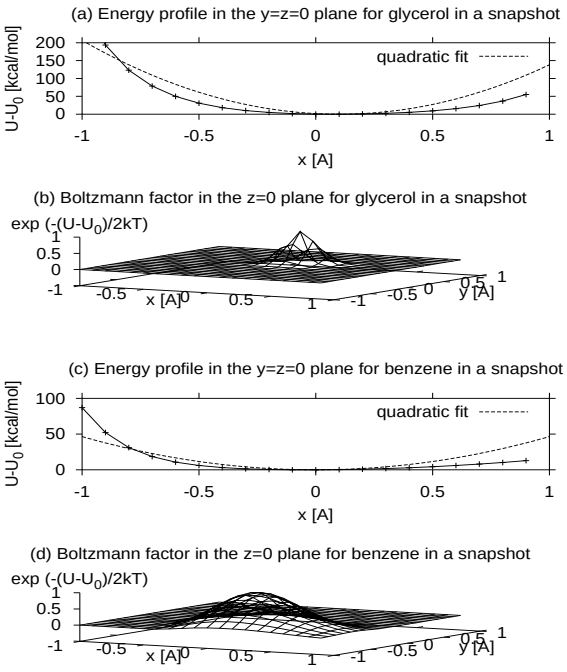

FIG. 1. Illustrations of the restriction of the motion of glycerol (a, b) and benzene (c, d) in the field of the water molecules: (a) One-dimensional energy profile in the plane y=z=0, for displacements of the glycerol molecule along the x-axis. A quadratic fit to the data illustrates the anharmonic behavior of the energy. (b) Two-dimensional profile of the Boltzmann factor $\exp[-\frac{\Delta U(x,y,0,\{\mathbf{q_w}\})}{kT}]$ for glycerol moving along x and y in the plane z=0. (c) and (d) are the corresponding plots for benzene.

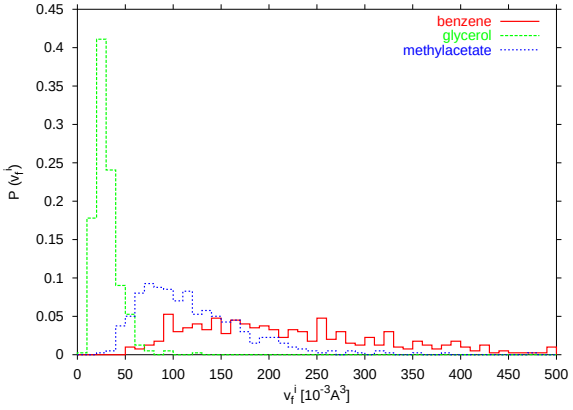

FIG. 2. Normalized probability distributions of free volume for three solutes of similar size in water : Glycerol (green), Methylacetate (blue) and Benzene (red). v_f^i data are binned in 10 Å^3 increments to generate the histograms.

the benzene, because of the presence of both hydrophobic (methyl) and hydrophilic (acetate) groups.

The average free volumes are computed as $\bar{v_f} = \frac{1}{N}\sum_{i=1}^{N} v_f^i$ (where N = number of snapshots) with a standard deviation $\sigma(v_f) = (\frac{1}{N}\sum_{i=1}^{N}(v_f^i - \bar{v_f})^2)^{1/2}$, as summarized in Table I. For each solute considered in this study, the free volume is very small compared to the size of the solute itself, because it corresponds to the volume defined by the movement of the *center of mass* of the solute molecule. The large standard deviations could be thought to arise from using a limited number of steps in the simulations. However, the histograms after 400 ps and 800 ps are very similar (data not shown), indicating that these deviations are a reflection of the local variations in density in the liquid state. Moreover, in the next section we perform similar calculations for pure water, for which we observe the same phenomenon.

In Table I we also report free volumes $v_{f,mf}$ calculated with potential energy profiles averaged over all the snapshots. These calculations systematically underestimate the average free volume. Two problems contribute to this effect, as seen on Fig. 1: (1) the energy profile of the different snapshot are not necessarily centered on the equilibrium position of the solute molecule, due to the kinetic terms; (2) since at the edges the energy rises very rapidly, energy averages are highly weighted in favor of the small volumes. Therefore we will use $\bar{v_f}$ instead of $v_{f,mf}$ to calculate the translational entropies.

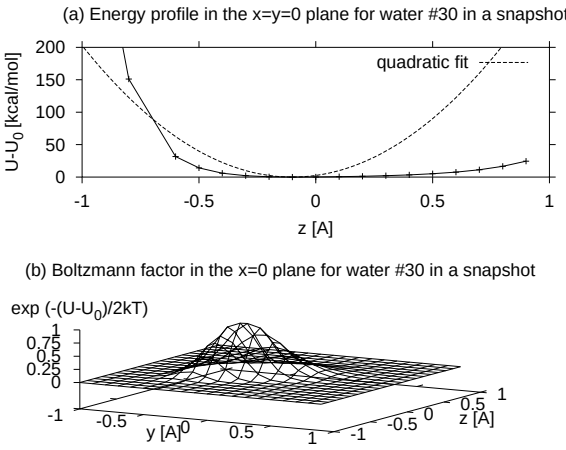

FIG. 3. Illustrations of the restriction of the motion of a water molecule in the field of the other molecules: (a) One-dimensional energy profile in the plane y=z=0, for displacements of a water molecule (number #30) along the x-axis. A quadratic fit to the data illustrates the anharmonic behavior of the energy. (b) Two-dimensional profile of the Boltzmann factor $\exp[-\frac{\Delta U(x^*,y^*,z^*,\{\mathbf{q_{w\neq w^*}}\})}{2kT}]$ in the plane z=0, for the same water molecule.

B The case of pure water

CM calculations for pure water require a slight adaptation of Eq. 6. In this case, the solute itself is a water molecule (denoted by '$*$' in the following), and the assumption (used to write Eq. 2) that the solutes do not interact with each other (ideal solution) is violated. Therefore, we need to take these interactions into account in the potential term of the Hamiltonian. The derivation in the appendix leads to the free volume of a specific water molecule (labeled '$*$')

$$v_{f,w^*} = \int_{\frac{V}{N_w}} e^{-\frac{\beta}{2}U(\mathbf{q_{w^*}},\{\mathbf{q_{N_{w\neq w^*}}}\})} d\mathbf{q_{w^*}} \qquad (10)$$

and the translational entropy of the whole water box is

$$S_{w,CM}^{trsl} = k(N_w \ln \frac{v_{f,w}}{\Lambda_w^3} + \frac{5}{2}N_w) \qquad (11)$$

The energy profiles of Fig. 1 (glycerol or benzene in water) and Fig. 3 (pure water) reveal that the range of motion of the center of mass of a molecule of glycerol or benzene in water is similar to that of a water molecule (less than 1 Å in each direction). This feature is not necessarily intuitive, considering the difference of size between these solutes and water. It emphasizes once more that restriction of motion of the *center of mass* of the molecule depends on the interactions of the molecule with its neighbors more than on the size of the molecule itself. Nevertheless, motion of a bigger molecule requires a coordinated displacement between more water molecules, which is less probable and therefore likely to reduce the free volume, as will be shown later in this study.

As in the previous section, we calculated the free volume for each snapshot, v_{f,w^*} and report in Fig. 4 the distribution of v_{f,w^*} for 4 different water molecules (labeled w1, w2, w3 and w4, respectively). The histograms for w1, w2 and w3 were calculated from the 800 ps simulation that used adjustment of the kinetic energy [23] to maintain the temperature at 300K, whereas that for w4 was calculated from a separate, independent 800 ps simulation that used a Nose-Hoover thermostat [24, 25].

The peak of Fig. 3br (pure water) is larger than that of Fig. 1b (glycerol in water) but narrower than that of Fig. 1d (benzene in water), which is again consistent with the ranking of the polar surface areas of glycerol, water and benzene (see Table I).

All four histograms in Fig. 4 have similar widths and are skewed to the right (i.e., tail at larger values of v_f^i). The sharp cutoff at low v_f^i corresponds to densely packed regions and the tail at high v_f^i to low-density regions. As in the previous section, the width of the distribution reflects the local density variations in liquid water [27].

The average value of the free volume from the combination of these four histograms is 0.099 Å^3, with a standard deviation of 0.088 Å^3. With this value, the translational entropy of water

Table I. Comparison of free volumes (in Å^3) for various solutes in water, for a molecule of pure water, and for benzene in the TCM T4-lysozyme. $\bar{v_f}$ is the average free volume over all snapshots and $\sigma(v_f)$ its standard deviation. $v_{f,mf}$ is the free volume calculated after averaging the potential, as explained in the text. ASA$_T$ and ASA$_P$ are the total accessible surface area and the polar accessible surface area of the molecule, respectively.

Molecule i (by decreasing size)	ASA$_T$ [Å^2]	ASA$_P$ [Å^2]	v_f [Å^3]	
			$v_{f,mf}$	$\bar{v_f} \pm \sigma(v_f)$
4-methly, 1-hydroxyl naphthalene	322.8	47.17	0.035	0.053 ± 0.030
naphthylhydroquinone	314.4	94.71	0.021	0.030 ± 0.013
naphthylquinone	312.8	90.96	0.032	0.049 ± 0.024
naphthalene	286.7	0.000	0.066	0.146 ± 0.084
quinone	255.5	102.2	0.035	0.059 ± 0.031
glycerol	238.2	154.6	0.024	0.030 ± 0.013
benzene	225.0	0.000	0.101	0.245 ± 0.147
methylacetate	218.8	64.46	0.070	0.120 ± 0.059
methanol	148.9	61.54	0.078	0.129 ± 0.073
pure water	113.1	113.1	0.049	0.099 ± 0.088
benzene in TCM T4-Lysozyme			0.053	0.095 ± 0.036

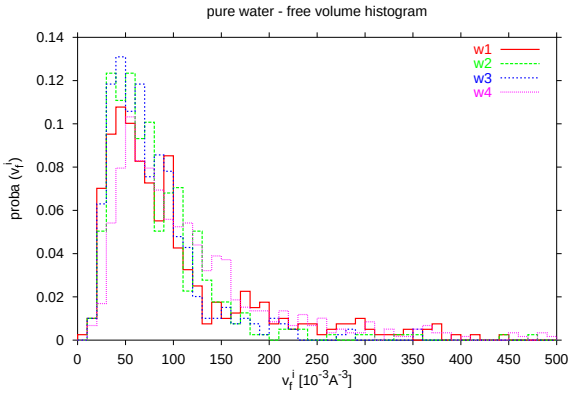

FIG. 4. Normalized probability distributions of free volumes for 4 different water molecules, calculated for snapshots taken every 2 ps during the 800 ps simulation. v_f^i data are binned in 10 Å^3 increments to generate the histograms.

is $S_{w,CM}^{trsl} = 8.90 \pm 1.76\, e.u.$ (the error on the entropy is given by $\sigma(S) = R\frac{\sigma(v_f)}{\bar{v_f}}$). The ST estimate (Eq. 5) at the density of liquid water, gives $S_{w,ST}^{trsl} = 20.3\, e.u.$. To compare the CM and ST estimates with the experimental value of the *total* entropy of liquid water at 298 K [28] :

$$S_{w,\mathrm{EXP}}^{total} = 16.7\, e.u.$$

we need to estimate the *rotational* entropy of water (S_w^{rot}), assuming that the vibrational and other contributions are comparatively small at this temperature [29]. We can approximately determine lower and higher bounds of S_w^{rot} from the entropy of melting and from gas-phase values, respectively. The experimental value [28] of the entropy of melting of ice to water is $S_{melt,273K} = 5.258\, e.u.$. Corrected to 298K using the heat capacity (C_p) of ice : $S_{melt,298K} = 6.835\, e.u.$. Assuming the two main contributions to $S_{melt,298K}$ are the rotational and communal entropies [30], we obtain a value for the rotational contribution to the entropy S_w^{rot} of $4.85\, e.u.$. This value is in reality the difference between the rotational entropy in water and the librational entropy in ice and should thus be considered a lower limit (i.e., $S_{w,298K}^{rot} \geq 4.85\, e.u.$). The gas-phase rotational entropy for water [14] gives the upper limit ($S_{w,298K}^{rot} \leq 10.4\, e.u.$). Adding these rotational entropy bounds to the CM estimate ($S_{w,CM}^{trsl} \approx 8.90\, e.u.$), the limiting values of the total entropy are:

$$13.75\, e.u. \leq S_{w,\mathrm{CM}}^{total} \leq 18.4\, e.u.$$

The same procedure with the ST estimate ($S_{w,ST}^{trsl} \approx 20.3\, e.u.$) leads to

$$25.75\, e.u. \leq S_{w,\mathrm{ST}}^{total} \leq 30.8\, e.u.$$

These limiting values are compared with the experimental information in Table II.

The CM and ST estimates can also be compared to the experimental entropy of vaporization [28], at a standard state of 1 atm and 298K :

$$\Delta S_{vap,\mathrm{EXP}}^{total} = 28.29\, e.u.$$

We assume the main contributions to the entropy of vaporization are translational ($\Delta S_{vap}^{trsl} = S_{gas}^{trsl} - S_{liq}^{trsl} = R\ln(v_{f,gas}/v_{f,liq})$) and rotational ($\Delta S_{vap}^{rot} = S_{gas}^{rot} - S_{liq}^{rot}$) and use the same approximations as before for the rotational contributions: $4.85\, e.u. \leq S_{liq}^{rot} \leq 10.4\, e.u. = S_{gas}^{rot}$. For the translational contribution, S_{gas}^{trsl} can be estimated from Eq. 5 (with $V/N = v_{f,gas} = 24,400\, cm^3/mol$, at 1 atm and 298 K). Using the CM model, S_{liq}^{trsl} can be estimated from Eq. 7 ($v_f = 0.099 Å^3/molecule$), which leads to

$\Delta S^{trsl}_{vap,CM} = 21.02\, e.u.$ Using the ST model, S^{trsl}_{liq} can be estimated from Eq. 5, at the density of liquid water, which leads to $\Delta S^{trsl}_{vap,ST} = 14.27\, e.u.$.

Combining the rotational and translational estimates, we get limiting values of the total entropy of vaporization for the CM model:

$$21.02\, e.u. \leq \Delta S^{total}_{vap,\mathrm{CM}} \leq 26.57\, e.u.$$

and for the ST model:

$$14.27\, e.u. \leq \Delta S^{total}_{vap,\mathrm{ST}} \leq 20.16\, e.u.$$

The comparison between these values and the experimental one is summarized in Table II.

This study reveals that CM provides a much better approximation of the translational entropy of water than ST. This is not very surprising, considering that ST is derived from ideal gas statistical mechanics, neglecting the interactions between molecules, whereas CM takes the intermolecular interaction explicitely into account in calculating the free volume, as seen in Eq. 7.

Table II. Comparison of the experimental values (EXP) with our Cell Model (CM) and with the Sackur-Tetrode (ST) estimates. S^{total}_{w} is the entropy of liquid water at 298K and ΔS^{total}_{vap} is the entropy of vaporization of water at 298K. The arrows indicate boundaries calculated as explained in the text. Values are in entropic units (e.u. = $\frac{cal}{mol\ K}$).

	ST	CM	EXP
S^{total}_{w}	25.15 → 30.8	13.75 → 18.4	16.7
ΔS^{total}_{vap}	14.27 → 20.16	21.02 → 26.57	28.29

Some of the discrepancies between CM and the experimental values might be a consequence of the assumption that each cell is occupied by one molecule. This overlooks the regions of low density that exist locally in water, as indicated by the tail on the histogram of Fig. 4. Some modifications of CM, notably the hole theories, take into account the possibility of unoccupied cells, but these lead to many practical complications without significantly improving the results [14, 30].

C Benzene binding to the TCM T4-lysozyme

As mentioned above, the TCM T4-lysozyme was designed to accomodate a benzene molecule in a cavity created by three mutations : L99A, C54T and C97A.

1 Normal Mode Analysis

Fig. 5 presents the energy profile for small displacements of the benzene molecule in the cavity of the TCM T4-lysozyme, along each of the ten lowest-frequency modes. Modes 1 to 6 exhibit frequencies about 4 times lower than the next modes and their energy profile has a relatively flat minimum corresponding to anharmonic translational/rotational motions (similar to a particle-in-a-box motion). Modes 7 to 10 correspond to low frequency internal modes, with a clear harmonic behavior. This analysis shows that the motion of the benzene within the binding site of the TCM T4-lysozyme can be best described as a combination of translations.

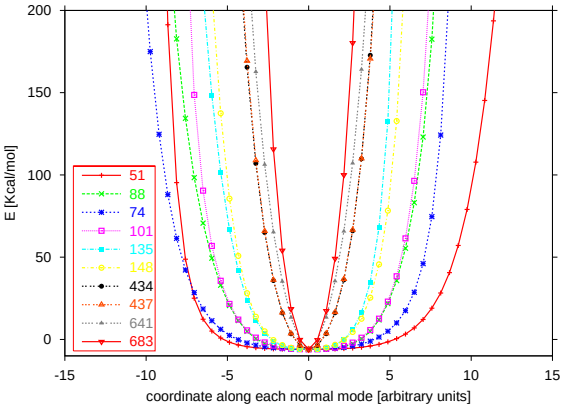

FIG. 5. Energy profile along the ten lowest-frequency modes. The frequency of each mode (in cm^{-1}) is listed in the figure. Modes 1 to 6 (51 to 148 cm^{-1}) exhibit a relatively flat energy minimum corresponding to translations/rotations, whereas modes 7 to 10 (434 to 683 cm^{-1}) correspond to higher frequency harmonic motions.

2 Free volume

Fig. 6 shows a histogram of the free volume of benzene in the hydrophobic cavity of TCM T4-lysozyme. A comparison of Fig. 6 and Fig. 2 indicates that the average free volume of benzene in the TCM T4-lysozyme is much smaller than that of benzene in water (see Table I). Moreover, the general shape of the histogram in Fig. 6 is similar to that of glycerol in water. This reflects the tight encapsulation of the benzene by the hydrophobic cavity as well as the higher density of the protein interior (about 1.4 times that of water [31]).

Simple concentration estimates (such as the "cratic entropy"[2, 12, 32]) as well as gas phase estimates (Sackur-Tetrode equation) would disregard the difference in free volume between benzene in water and benzene in the TCM T4-lysozyme, as mentioned in the section II B. Using Eq. 8 and considering that the free volume of the TCM T4-lysozyme is similar to that of the complex, we get $\Delta S^{trsl}_{CM} \approx -11.78\, e.u.$.

Hermans and Wang [20] estimated the rms displacements of the benzene in the binding pocket using MD simulations and obtained a value of 0.13 Å for the motion perpendicular to the benzene ring, and 0.3 Å in two orthogonal directions in the plane of the ring, which corresponds to a volume of about

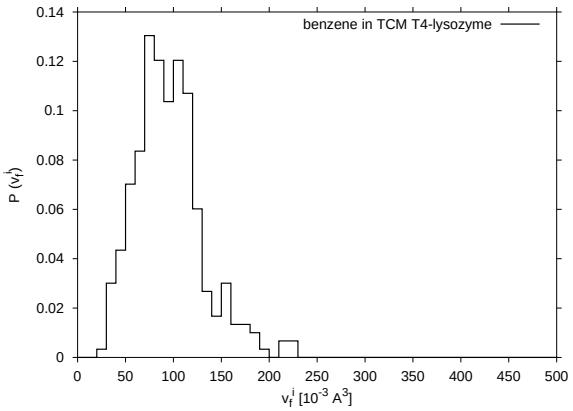

FIG. 6. Normalized probability distributions of free volume for benzene in TCM T4-lysozyme. v_f^i data are binned in 10 Å^3 increments to generate the histogram.

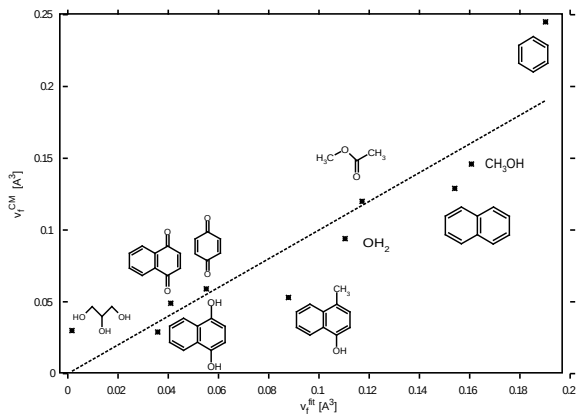

FIG. 7. Free volume calculated with the Cell Model (v_f^{CM}) vs. free volume calculated from a correlation with total and polar surface areas (v_f^{fit}) for the various solutes used in this study as well as for pure water. The perfect fit is given by the line ($v_f^{CM} = v_f^{fit}$).

$0.26 \times 0.6 \times 0.6 = 0.0936$ Å^3, in excellent agreement with our free volume calculations.

D Parameterization

The results obtained for the free volumes shown in Fig. 2 suggest some degree of correlation between the free volume and the hydrogen-bonding ability or polarity of the solute. To quantify this, we calculated the total (ASA_T) and polar (ASA_P) surface area of each solute listed in Table I (including the case of pure water), using an algorithm originally developped by Lee and Richards [33]. The free volume calculated with the cell model (v_f^{CM}) was fitted to a linear combination of polar and total surface areas of the molecule :

$$v_f^{fit} = a\,\mathrm{ASA}_P + b\,\mathrm{ASA}_T + c$$

with good agreement (Fig. 7). The value of the fitting parameters are :

$$a = -1.18\,10^{-3}\,\text{Å} \qquad b = -4.78\,10^{-4}\,\text{Å} \qquad c = 0.2978\,\text{Å}^3$$

This indicates that for the range of solutes considered in this study, the maximum free volume is c=0.2978 Å^3. It also indicates that the free volume of a molecule is more influenced by its polarity than by its size (i.e. $a > b$). This correlation applies to molecules with ASA_T smaller than about 600 Å^2. Larger molecules, including proteins, are expected to have a free volume of approximately 0.03 Å^3 or smaller.

Even though more data on several other solutes are necessary to establish this correlation more precisely, this promising result opens the possibility to obtain estimates of the translational entropy of a solute in water bypassing the molecular dynamics simulations.

IV. SUMMARY AND CONCLUSION

Theoretical estimation of the loss of entropy upon molecular association (ΔS) is essential for rationalizing binding affinities as well as predicting entropic contribution to enzyme catalysis [1–9]. The main difficulty in estimating ΔS from first principles resides in the intricacy of the intermolecular interactions in solution.

In this study we developed and analysed a framework, based on cell models (CM) of liquids, to evaluate the loss of translational entropy ΔS^{trsl} [2]. The CM, as implemented in this study, takes the intermolecular potential explicitly into account using information gathered during molecular dynamics (MD) simulations in solution. In the CM framework, the entropy of a solute in solution is related to the volume in which the center of mass is allowed to move without the molecule colliding with the neighboring, fixed, water molecules. This so-called "free volume" was calculated for various solutes in water, as well as for pure water.

The estimated free volume seems to vary significantly from compound to compound (Table I), ranging from 0.029 Å^3 (for naphthylhydroxyquinone) to 0.245 Å^3 (for benzene). We observed a correlation between this free volume and a linear combination of the total and polar surface area of the solute, suggesting that the free volume (and therefore the translational entropy) of a molecule can be estimated based solely on its surface area (total and polar), without having to carry out costly MD simulations.

The Sackur-Tetrode (ST) equation, based on gas-phase statistical mechanics, is sometimes used for proteins to circumvent the difficulties of liquid-state statistical mechanics [3, 13]. The CM predictions for pure water agree with the experimental data, much better than the ST equation, suggest-

ing that CM rather than ST should be used to estimate the translational entropy loss in solution. We also applied the CM to the estimation of ΔS^{trsl} upon binding of benzene to the TCM T4-lysozyme. The results are in excellent agreement with previous estimates [20], demonstrating the applicability of the CM.

The CM also provides a theoretical explanation of the simple concentration estimates based on the "cratic entropy" [12, 32]: these neglect the difference in free volume between a ligand in water and a ligand in the binding site. For the case of benzene binding to the TCM T4-lysozyme ($v_{f,\mathrm{benzene\,in\,water}} = 0.245\text{Å}^3$; $v_{f,\mathrm{benzene\,in\,lysozyme}} = 0.095\text{Å}^3$), this leads to an error in the entropy of approximately $1.88\,e.u.$.

We therefore propose the CM as a simple method to evaluate the loss of translational entropy upon molecular association. An extension of this model for the other components (e.g., rotational) of this entropy loss will be considered in a future publication.

ACKNOWLEDGMENTS

We acknowledge the National Cancer Institute for allocation of computing time and support staff at the Frederick Biomedical Supercomputing Center of the Frederick Cancer Research and Development Center. This work was supported by grant GM51362 from NIGMS.

APPENDIX A: PARTICULAR CASE : PURE WATER

This appendix contains derivations for the calculation of the free volume and translational entropy of pure water, which can be considered a particular case of a solute ξ in water, in which the solute itself is a water molecule. However, as mentioned in the text, in this case the assumption (used to write Eq. 2) that the "solutes" do not interact with each other (ideal solution) cannot be used. Therefore, two modifications are necessary: 1) take these interactions into account in the potential term of the Hamiltonian and 2) use the partition function of the system Q_w instead of the partition function of one molecule q_w (used in Eq. 2). For N_w molecules of water in a volume V at a temperature T, the translational entropy of water is given by (cf. Eq. 3):

$$S_w^{trsl} = k(\ln Q_w + T\frac{\delta \ln Q_w}{\delta T}|_V) \tag{A1}$$

where the partition function of the whole system is

$$Q_w = \frac{1}{h^{3N_w} N_w!}\int\int e^{-\beta H(\mathbf{p_w},\mathbf{q_w})} d\mathbf{p_w} d\mathbf{q_w} \tag{A2}$$

and the Hamiltonian:

$$H(\mathbf{p_w},\mathbf{q_w}) = \frac{\mathbf{p_w}^2}{2m_\xi} + U(\mathbf{q_w}) \tag{A3}$$

Integration of Eq. A2 over the momenta yields

$$Q = \frac{1}{\Lambda_w^{3N_w} N_w!} Z \tag{A4}$$

Where $\Lambda_w = h/\sqrt{2\pi m_w kT}$ is the De Broglie wavelength of a water molecule [14] and we introduced the configurational integral Z:

$$Z = \int e^{-\beta U(\mathbf{q_w})} d\mathbf{q_w} \tag{A5}$$

By analogy with the gas phase entropy, we write the liquid entropy as

$$S_{w,liq}^{trsl} = k(N_w \ln \frac{v_{f,w}}{\Lambda_w^3} + \frac{5}{2}N_w) \tag{A6}$$

This defines the free volume as

$$Z = (Nv_f)_N = \int_1 \ldots \int_N e^{-\beta U(\mathbf{q_1}\ldots\mathbf{q_N})} d\mathbf{q_1}\ldots d\mathbf{q_N} \tag{A7}$$

Dividing the space in N_w cells for each of the integrals: $\int = \int_{c_1} + \ldots + \int_{c_{N_w}}$, with one water molecule in each cell.

$$Z = \sum_{\alpha_1=c_1}^{c_N} \ldots \sum_{\alpha_N=c_1}^{c_N} \int_{\alpha_1} \ldots \int_{\alpha_N} e^{-\beta U(\mathbf{1}\ldots\mathbf{q_N})} d\mathbf{q_1}\ldots d\mathbf{q_N} \tag{A8}$$

As there are N^N ways to place N molecules into N cells, we get

$$Z = N^N \int_{c_1} \ldots \int_{c_N} e^{-\beta U(\mathbf{q_1}\ldots\mathbf{q_N})} d\mathbf{q_1}\ldots d\mathbf{q_N} \tag{A9}$$

Assuming pairwise additivity of the potential:

$$U(\mathbf{q_1}\ldots\mathbf{q_N}) = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} U_{ij}(\mathbf{q_i},\mathbf{q_j}) \tag{A10}$$

To reduce the N-dimensional integral to one dimension, we consider that each molecule is moving in the mean field created by the others. Therefore we are reduced to the evaluation of the integral for one given water molecule (labeled '$*$') located in the cell 'c_*'

$$Z = (Nv_f)^N = N^N(\int_{c_*} e^{-\frac{\beta}{2}U(\mathbf{q}_*,\{\mathbf{q}\neq\mathbf{q}_*\})} d\mathbf{q}_*)^{\mathbf{N}} \tag{A11}$$

which gives the free volume of a water molecule:

$$v_{f,w_*} = \int_{c_*} e^{-\frac{\beta}{2}U(\mathbf{q}_*,\{\mathbf{q}\neq\mathbf{q}_*\})} d\mathbf{q}_* \tag{A12}$$

REFERENCES

(1) Gilson, M. K., Given, J. A., Bush, B. L., and McCammon, J. A. The statistical-thermodynamic basis for computation of binding affinities: a critical review. Biophys. J. 72(3):1047–69, 1997.

(2) Amzel, L. M. Loss of translational entropy in binding, folding, and catalysis. Proteins 28(2):144–149, 1997.

(3) Brady, G. P. and Sharp, K. A. Entropy in protein folding and in protein-protein interactions. Curr. Opin. Struct. Biol. 7(2):215–221, 1997.

(4) Tamura, A. and Privalov, P. The entropy cost of protein association. J. Mol. Biol. 273:1048–1060, 1997.

(5) Ben-Tal, N., Honig, B., Bagdassarian, C. K., and Ben-Shaul, A. Association entropy in adsorption processes. Biophys. J. 79(3):1180–1187, 2000.

(6) Page, M. I. and Jencks, W. P. Entropic contributions to rate accelerations in enzymic and intramolecular reactions and the chelate effect. Proc. Natl. Acad. Sci. USA 68(8):1678–1683, 1971.

(7) Lightstone, F. C. and Bruice, T. C. Ground state conformations and entropic and enthalpic factors in the efficiency of intramolecular and enzymatic reactions. J. Am. Chem. Soc. 118:2595–2605, 1996.

(8) Villà, J., Strajbl, M., Glennon, T. M., Sham, Y. Y., Chu, Z. T., and Warshel, A. How important are entropic contributions to enzyme catalysis ? Proc. Natl. Acad. Sci. USA 97(22):11899–11904, 2000.

(9) Luo, H. and Sharp, K. On the calculation of absolute macromolecular binding energies. Proc. Natl. Acad. Sci. USA 99(16):10399–10404, 2002.

(10) Steinberg, I. Z. and Scheraga, H. A. Entropy changes accompanying association reactions of proteins. J. Biol. Chem. 238(1):172–181, 1963.

(11) Janin, J. For Guldberg and Waage, with love and cratic entropy. Proteins 24:i–ii, 1996.

(12) Janin, J. Elusive affinities. Proteins 21:30–39, 1995.

(13) Finkelstein, A. V. and Janin, J. The price of lost freedom: Entropy of bimolecular complex formation. Protein Engineering 3(1):1–3, 1989.

(14) Hill, T. L. An introduction to statistical thermodynamics. Dover Publications Inc., New York, 1960.

(15) Hirschfelder, J. O., Stevenson, D., and Eyring, H. A theory of liquid structure. J. Chem. Phys. 57:896–912, 1937.

(16) Yu, Y. B., Privalov, P. L., and Hodges, R. S. Contribution of translational and rotational motions to molecular association in aqueous solution. Biophys. J. 81(3):1632–1642, 2001.

(17) Eyring, H. and Hirschfelder, J. O. The theory of the liquid state. J. Chem. Phys. 41:249–257, 1937.

(18) Kirkwood, J. G. Critique of the free volume of the liquid state. J. Chem. Phys. 18(3):380–382, 1950.

(19) Luo, R. and Gilson, M. K. Synthetic adenine receptors: Direct calculation of binding affinity and entropy. J. Am. Chem. Soc. 122(12):2934–2937, 2000.

(20) Hermans, J. and Wang, L. Inclusion of loss of translational and rotational freedom in theoretical estimates of free energies of binding. Application to a complex of benzene and mutant T4 lysozyme. J. Am. Chem. Soc. 119:2707–2714, 1997.

(21) Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., and Karplus, M. CHARMm - a program for macromolecular energy, minimization, and dynamics calculations. J. Comput. Chem. 4(2):187–217, 1983.

(22) Ryckaert, J. P., Cicotti, G., and Berendsen, H. J. C. Numerical integration of Cartesian equations of motion of a system with constraints: Molecular dynamics of n-alkanes. J. Comput. Phys. 23:327–341, 1977.

(23) Berendsen, H. J. C., Postma, J. P. M., VanGunsteren, W. F., Dinola, A., and Haak, J. R. Molecular-dynamics with coupling to an external bath. J. Chem. Phys. 81(8):3684–3690, 1984.

(24) Hoover, W. G. Canonical dynamics - Equilibrium phase-space distributions. Phys. Rev. A 31(3):1695–1697, 1985.

(25) Nose, S. A unified formulation of the constant temperature molecular-dynamics methods. J. Chem. Phys. 81(1):511–519, 1984.

(26) Eriksson, A. E., Baase, W. A., Wozniak, J. A., and Matthews, B. W. A cavity-containing mutant of T4 lysozyme is stabilized by buried benzene. Nature 355(6358):371–373, 1992.

(27) Eisenberg, D. and Kauzman, W. The structure and properties of water. Oxford University Press, London, 1969.

(28) Cox, J. D., Wagman, D. D., and Medvedev, V. A. CODATA Key Values for Thermodynamics. Hemisphere Publishing Corp., New York, 1989.

(29) Dunitz, J. D. The entropic cost of bound water in crystals and biomolecules. Science 264(5159):670–670, 1994.

(30) Hirschfelder, J. O., Curtiss, C. F., and Bird, R. Molecular theory of gases and liquids. John Wiley and Sons Inc., New York, 1964.

(31) Richards, F. M. The interpretation of protein structures: Total volume, group volume distributions and packing density. J. Mol. Biol. 82(1):1–14, 1974.

(32) Murphy, K. P., Xie, D., Thompson, K. S., Amzel, L. M., and Freire, E. Entropy in biological processes: Estimation of translational entropy loss. Proteins 18(1):63–67, 1994.

(33) Lee, B. and Richards, F. M. The interpretation of protein structures: Estimation of static accessibility. J. Mol. Biol. 55(3):379–400, 1971.