ISSN 0006-2979, Biochemistry (Moscow), 2007, Vol. 72, No. 7, pp. 785-792. © Pleiades Publishing, Ltd., 2007.
Published in Russian in Biokhimiya, 2007, Vol. 72, No. 7, pp. 963-973.

Computation of Entropy Contribution to Protein–Ligand Binding Free Energy

F. V. Grigoriev[1]*, S. V. Luschekina[2], A. N. Romanov[1], V. B. Sulimov[1], and E. A. Nikitina[3]

[1]*Research Computing Center of Lomonosov Moscow State University, 119992 Moscow, Russia; E-mail: fedor.grigoriev@gmail.com*
[2]*Orekhovich Institute of Biomedical Chemistry, ul. Pogodinskaya 10, 119121 Moscow, Russia; E-mail: sofia.luschekina@gmail.com*
[3]*Institute of Applied Mechanics, Russian Academy of Sciences, Leninsky pr. 32a, 119991 Moscow, Russia; E-mail: nikitina.ekaterina@gmail.com*

Received December 20, 2006
Revision received April 26, 2007

Abstract—The entropy contribution ΔS to protein–ligand binding free energy is studied for nine protein–lipid complexes. The entropy effect from the loss of the translational/rotational degrees of freedom (ΔS_{tr}) is calculated using the ideal gas approach. The change in the vibrational entropy (ΔS_{vib}) is calculated using the effective quantum oscillator approach with frequencies derived from the coordinate covariance matrix, so the inharmonic effects are taken into account. The change in the entropy of solvation (ΔS_{solv}) is considered using the binomial cell model (developed by the authors) for the hydrophobic effect. The entropy contribution from loss of conformations that are available for the free ligand (ΔS_{conf}) is also estimated. It is revealed that the negative in view of binding term ΔS_{tr} is only partly compensated by increasing of ΔS_{vib}, so $T(\Delta S_{tr} + \Delta S_{vib} + \Delta S_{conf}) < 0$ for all complexes under investigation, but taking into account ΔS_{solv} leads to significantly increased ΔS. For all complexes except biotin–streptavidin, the results are found to be in reasonable agreement with experimental data.

DOI: 10.1134/S0006297907070140

Key words: protein–ligand complex, binding free energy, covariance matrix, vibration entropy

One of the most important factors limiting the formation of intermolecular protein–ligand (PL) complexes is the restriction of rotational, translational, and internal degrees of freedom both for ligand and protein. The corresponding contribution to the PL binding free energy ΔG_b is mainly due to the change in configurational entropy ΔS_{config}. Computation of ΔS_{config} is a complicated task because the full investigation of available configurational space is needed. Different simplified approaches are used for computation of ΔS_{config} for real systems.

The configurational integral is described as a sum of contributions from lowest states for computation of change in ΔS_{config} [1]. For computation of the contribution of every state, harmonic analysis was done at a corresponding local minimum [2]. The configurational space available for the ligand is estimated based on analysis of the trajectory, which is obtained from molecular dynamic (MD) simulation [3]. The ligand is described as an absolutely rigid body, so the contribution to ΔS_{config} from internal degrees of freedom is neglected. Different schemes for computation of ΔS_{config} using MD simulation were considered in [4]. The contribution from rotational and translational degrees of freedom was estimated from available bonded state configurational space, which is calculated from the quasi-harmonic analysis of MD trajectory.

The ΔS_{config} has been evaluated for insulin dimer complex using decomposition of ΔS_{config} into rotational, translational (ΔS_{tr}), and vibrational (ΔS_{vib}) terms [5]. For ΔS_{tr} the ideal gas approach was used, and for ΔS_{vib} the quantum oscillator approach with frequencies computed from harmonic analysis of protein, ligand, and PL complex was used. It was found that the increase in ΔS_{vib} from the additional vibrational degrees of freedom was insufficient for compensation of ΔS_{tr}, and the total change is $T\Delta S_{config} = -20.1$ kcal/mol.

The authors of [6] used a decomposition similar to that given in [5] for computation of ΔG_b for the

Abbreviations: BCM) Binomial Cell model; CM) covariance matrix; GB) Generalized Born approach; MD) molecular dynamic.

* To whom correspondence should be addressed.

trypsin–benzamidine complex. Reasonable agreement of the calculated and experimental values of ΔG_b was achieved [6], though the values of the empirical parameters used for the computation are questionable and require additional arguments.

Now one of the most intensively developed methods for computation of ΔS_{config} is based on analysis of the covariance matrix (CM) obtained from MD simulation. The CM approach was first used in [7], where the entropy change for the two states of the system was calculated as $\Delta S = 0.5k_B\ln[\det(\sigma_A)/\det(\sigma_B)]$, where $\det(\sigma_A)$ and $\det(\sigma_B)$ are determinants of the CM for states A and B, respectively, and k_B is the Boltzmann constant. The quantum correction to ΔS_{config} was suggested in [8] for the system of decaalanine which contains 53 atoms. In [9], the CM approach with the quantum correction was applied to a system of 256 argon atoms. Also, it was demonstrated that the value of ΔS_{config} calculated in this way converges with increasing time of MD modeling. In [10] a similar approach was used for several systems: a set of 100 independent oscillators, liquid argon, and peptide in solvent. In the first two cases, the entropy from CM was compared with the result of exact theoretical computation. It was demonstrated that the CM approach overestimates the entropy by up to 20%.

The MC approach has been used for biological systems (including PL complexes) [4, 11-15]. In work [11], four complexes were investigated: one of them was biotin–streptavidin, with $T\Delta S_{config} = -6.7$ kcal/mol, and the others were formed by relatively small molecules. The change in entropy calculated for biotin–streptavidin is close to –4.5 kcal/mol, which was reported in [16] for this complex with frequencies derived from harmonic analysis. In [12] ΔS_{config} was calculated for a heptapeptide in water without any restrictions on the coordinates of the atoms. To fully search the configurational space of this system, a MD trajectory with 200 nsec length was obtained and analyzed. The translational and rotational degrees of freedom of the heptapeptide as a whole were eliminated from the MC. The $T\Delta S_{config}$ of the heptapeptide increased on time of modeling, undergoing jumps of up to 2 kcal/mol, the last of them appearing at 180 nsec. These jumps are connected with change in conformation of the heptapeptide, which is confirmed by its time correlation. In [13] similar computations were made for α-lactalbumin. Because of the large size of this system the length of the MD trajectory was 2 nsec, and the $T\Delta S_{config}$ did not converge: during the last 100 psec it increased by 6 kcal/mol. In [14] the change in $T\Delta S_{config}$ was calculated for integrase with CD4 receptor. Only for relative small fragments of the complex (no more than 200 atoms) the convergence of $T\Delta S_{config}$ was achieved with CM obtained from the 16 nsec trajectory.

So the computation of ΔS_{config} using the CM for PL complexes without any restrictions to the internal degrees of freedom demands very large (tens of nanoseconds) trajectories, and even at this case the convergence of ΔS_{config} is questionable [12-14]. One of the possible ways to solve this problem is the separation in MD modeling of relative fast (bond and valence vibrations) and slow (torsions) degrees of freedom. In this case the computation of ΔS_{config} can be done similarly [1, 2]: first the search for conformations with lowest value of energy is made, and then for each of them the change in vibrational entropy ΔS_{vib} is calculated using the MC derived from MD modeling at fixed conformation. The full change in ΔS_{config} can be calculated as a sum of contributions ΔS_{vib} from different conformations with proper weights.

To use this scheme, it is necessary to develop the technique of computation of ΔS_{vib} using the CM approach with fixed conformation. In the present work, these computations are performed for the first time for nine PL complexes: eight of them with trypsin (five formed by benzamidine and its *p*-substituted analogs, three others by relatively large ligands) and biotin–streptavidin complex. The parameters for MD modeling, which provide convergence of ΔS_{vib} with increasing time of MD modeling, are defined. Contributions to ΔS_{config} from loss of translational and rotational degrees of freedom, conformational entropy, and also the entropic term caused by the effect of partial desolvation of protein and ligand in the bound state are estimated. For all of these complexes the calculated $T\Delta S_{calc}$ and experimental $T\Delta S_{exp}$ values are compared.

To consider the solvation effect, the Generalized Born (GB) approach is used [17]. Geometry optimization and MD modeling was performed using CHARMM [18], and for computation of interaction between atoms the MMFF94 [19] force field was used.

METHODS OF INVESTIGATION

Vibrational entropy ΔS_{vib} is calculated as the difference of the corresponding values for PL, protein, and ligand:

$$\Delta S_{vib} = S_{vib}(PL) - (S_{vib}(P) + S_{vib}(L)), \qquad (1)$$

where designations P, L, and PL correspond to protein, ligand, and the complex, respectively. Every term in Eq. (1) is calculated using the standard expression for the entropy of an ensemble of quantum oscillators [20] with frequencies evaluated from CM, which is defined as:

$$\sigma_{ij} = M^{1/2} < (q_i - <q_i>)(q_j - <q_j>) > M^{1/2}, \qquad (2)$$

where $M^{1/2}$ is a matrix of masses with rank $3N$ (N is number of atoms) in which diagonal elements are equal to masses of atoms and non-diagonal elements are equal to zero, and $<q_i>$ ($q = x, y, z$) are averaged over the MD trajectory coordinates of the i-th atom. Frequencies are calculated from the eigenvalues λ_i of MC as:

$$w_i = \sqrt{\frac{k_B T}{\lambda_i}}. \quad (3)$$

Harmonic restraints to torsion angles of the ligand were applied to eliminate the transition from the initial conformation to another during MD modeling. Two sets of the initial values of torsion angles for every ligand were defined by means of the geometric optimization of ligands in the bound and free states. Also, during MD modeling it is necessary to exclude translation and rotation of the ligand as a whole because the corresponding effect is incorporated into ΔS_{tr}. To do this, the velocity of the center of masses of ligand and the total angular moment were kept at zero during MD modeling. Harmonic restraints for the protein atoms relative to their initial positions were also applied.

Loss of conformational entropy ΔS_{conf} was calculated as in [21]:

$$T\Delta S_{conf} = -Nk_{rot}, \quad (4)$$

where $k_{rot} = 0.4$ kcal/mol is an empirical parameter chosen using results of computations of ΔS_{conf} for the set of PL complexes, and N is number of rotatable bonds. It is assumed in Eq. (4) that only one conformation is available for the ligand in the bound state. The corresponding inconsistence is incorporated into the empirical parameter k_{rot}.

So the procedure of computation of ΔS_{conf} consists of the following main steps:

– computation of ΔS_{tr};

– preparation of PL complexes, proteins, and ligands for MD modeling;

– geometric optimization of ligands in the free state; MD modeling of the ligands with values of torsions that correspond to the optimized geometries; computation of S_{vib} (L);

– geometric optimization of PL complexes and proteins; MD modeling of PL complexes and proteins with harmonic restraints described above;

– computation of S_{vib} (PL), S_{vib} (P), and ΔS_{vib} using Eq. (1);

– computation of $T\Delta S_{conf}$ using Eq. (4) and $T\Delta S_{config} = T\Delta S_{tr} + T\Delta S_{vib} + T\Delta S_{conf}$.

To compare calculated and experimental values of the change in entropy ΔS, it is necessary to estimate the contribution to ΔS from the effect of the partial desolvation of protein and ligand in the bound state. The computation of the corresponding term for protein–ligand binding free energy is usually carried out using its decomposition into polar and nonpolar contributions [22]. A similar approach can be used for entropy taking into account the standard expression $S = -\partial G/\partial T$.

The entropic component of the nonpolar term of desolvation free energy of ligands $T\Delta S_{nonpol}$ (L) was calculated based on the Binomial Cell model (BCM) [23, 24] that was developed for evaluation of the nonpolar term of the solvation free energy of organic compounds. In agreement with the BCM the entropy of formation in an aqueous cavity for solute is proportional to its volume V: $S_{nonpol} = \vartheta\xi R\rho V$, where $\vartheta = 1.3$ and $\xi = 0.73$ are dimensionless parameters of the model, ρ = 1/(volume per water molecule) = 1/30 Å^{-3}. For similar computation for the PL complexes and proteins, we took into account that the nonpolar term of solvation free energy of a macromolecule is equal to the product of its surface area and surface tension coefficient [25]. So for the change in the nonpolar entropic term the following equation was used:

$$T\Delta S_{nonpol} = T(\partial\sigma/\partial T)(A_{PL} - A_P) - \vartheta\xi RT\rho V_L =$$

$$= 0.069\ (A_P - A_{PL}) + 0.019\ V_L, \quad (5)$$

where $T\Delta S_{nonpol}$ is in kcal/mol, A_{PL} and A_P are surface areas of the PL complex and protein, respectively, in Å^2, V_L is the volume of the ligand in Å^3, $\partial\sigma/\partial T$ is the derivative of surface tension coefficient on temperature (calculated according to [26]).

The nonpolar term of solvation free energy ΔG_{pol} was considered in using the GB approach [17] during MD modeling. Using the initial Born model for a point charge, which is placed at the center of the dielectric cavity with fixed radius [22], the following expression can be derived for the entropic component of the free energy:

$$T\Delta S_{pol} = -T(\partial\Delta G_{pol}/\partial T) \cong$$

$$\cong -(T\Delta G_{pol}/\varepsilon^2)(\partial\varepsilon/\partial T) \cong 0.02\ \Delta G_{pol}, \quad (6)$$

where ε is dielectric constant for water ($\partial\varepsilon/\partial T$ is calculated according to [27]). With the assumption that the main contribution to $\partial\Delta G_{pol}/\partial T$ is caused by the temperature dependence of dielectric constant, the ratio (6) is true for the GB approach. We note that a similar approach was used in [28] for the estimation of the entropic component of the binding free energy for intermolecular complexes in water.

The total change in entropy caused by PL complex formation was calculated as:

$$T\Delta S_{calc} = T\Delta S_{config} + T\Delta S_{solv}, \quad (7)$$

where

$$T\Delta S_{solv} = T\Delta S_{nonpol} + T\Delta S_{pol}. \quad (8)$$

A decomposition similar to that in Eq. (7) was used in [1].

Details of MD modeling. The initial structures of the PL complexes were taken from the Protein Data Bank (PDB) [29] (determined by X-ray crystallography).

Hydrogen atoms were added using the Reduce program [30], which optimizes the positions of OH, SH, NH_3^+ fragments, hydrogen atoms of methyl groups, and Asn and Gln side-chain amides. After the addition of the hydrogen atoms, we used a procedure of truncation of the protein to extract only the active site that is important for formation of the PL complexes. Truncated structures with size ~400-500 atoms (with hydrogen) were made using an original program elaborated in our laboratory, which keeps all protein residues that have at least one heavy atom at a distance no more than a given cutoff radius R_{cut} from any heavy atom of the ligand in the active site. The value of R_{cut} was 4-5 Å and was determined so all fragments of the protein forming strong contact with ligand – salt bridges and hydrogen bonds – were kept as specific hydrophobic pockets.

The harmonic restraints to atoms were made using the CONS HARM FORCE 1.0 MASS command of CHARMM. This means that the term U_{harm} (in kcal/mol) is added to the potential energy of the system:

$$U_{harm} = \sum_{i=1}^{N} \frac{1}{2} k_{iharm} (x_i - x_{0i})^2,$$

where N is number of atoms; x_i, x_{0i} are current and initial values of coordinates, respectively; $k_{iharm} = 2A_i$, A_i is atomic mass of an element (for instance, $A_i = 12$ for carbon).

Optimization of the structure of the ligands was performed without any restraints, but during MD modeling restraints to torsion were used as described above. The selected torsions were restrained around their optimum values using CHARMM command CONS DIHE FORCE K_φ. This adds the term U_{harm} (in kcal/mol) to the potential energy of the system:

$$U_{harm1} = \sum_{i=1}^{N_1} \frac{1}{2} K_\varphi (\varphi_i - \varphi_{0i})^2,$$

where N_i is number of restraints on torsions, φ_i, φ_{0i} are current and initial values of torsions (in radians), and K_φ is force constant.

The structures and MD trajectories were visually analyzed using the VMD (Visual Molecular Dynamics) program (http://www.ks.uiuc.edu/Research/vmd/).

RESULTS AND DISCUSSION

Nine PL complexes were investigated. The first set contained five *p*-substituted benzamidines (Fig. 1) that form complexes with trypsin (the PDB code for trypsin–benzamidine complex is 3PTB). The others ligands with relative large molecular weights are shown in Fig. 2. For benzamidine the main binding motif is the strong electrostatic interaction between the positively charged amidine group of benzamidine and the negatively charged Asp189 and polar atoms of Gly219 and Ser190. As a result of this interaction, a salt bridge and two charged hydrogen bonds are formed. For *p*-substituted benzamidines the map of the hydrogen bonds is similar except for the case with R = $-CONH_2$, in which an additional hydrogen bond is formed. Detailed maps of the hydrogen bonds for the three large ligands to trypsin are shown and analyzed in [31, 32]. All these ligands retain strong contacts that are formed in the trypsin–benzamidine complex, and three additional hydrogen bonds are formed.

All hydrogen bonds are retained during molecular mechanic optimization, which was carried out following the scheme described above.

After optimization, the MD modeling was done under the following conditions: trajectory step size, 1 fsec; heating room temperature (298K) and equilibration were carried out for 29,000 and 40,000 steps, respectively, and coordinates of atoms were saved every 10 steps.

As discussed above, one of the key problems in the computation of ΔS_{vib} using MC is the length of the MD trajectory. The value of ΔS_{vib} will increase until the system visits all areas of the configurational space. Such behavior is shown for the trypsin–benzamidine complex (Fig. 3). The vibrational entropy for the complex (curve *1*) and for the protein (curve *2*) changed significantly during the first 200-300 psec, and after that they converged. The value $T\Delta S_{vib} = T(S_{vib}(PL) - S_{vib}(P) - S_{vib}(L))$ changes with time so that $T(S_{vib}(PL) - S_{vib}(P))$ (curve *3* on Fig. 3) because $TS_{vib}(L)$ for benzamidine converges during 5 psec. Similar behavior of $T\Delta S_{vib}$ with increasing length of trajectory was observed for all of the PL complexes. For relatively large ligands TS_{vib} (L) converges more slowly than for benzamidine and theirs *p*-substitute analogs, but in these cases the necessary length of trajectory is defined by the saturation of $T(S_{vib}(PL) - S_{vib}(P))$ and does not exceed 500 psec.

So using the harmonic restraints to the atoms of the protein and ligands by the above described scheme leads to a significant reduction in the length of the MD trajectory that is necessary to obtain the correct value of ΔS_{vib} with fixed conformation of protein and ligand. Using lengths of the MD trajectories similar to those used in

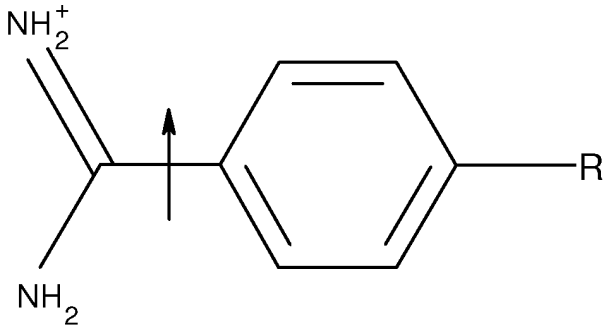

Fig. 1. Structural formulas of *p*-substituted benzamidines. R-groups in *p*-position: -H, $-NH_2$, $-CH_3$, $-OCH_3$, $-CONH_2$.

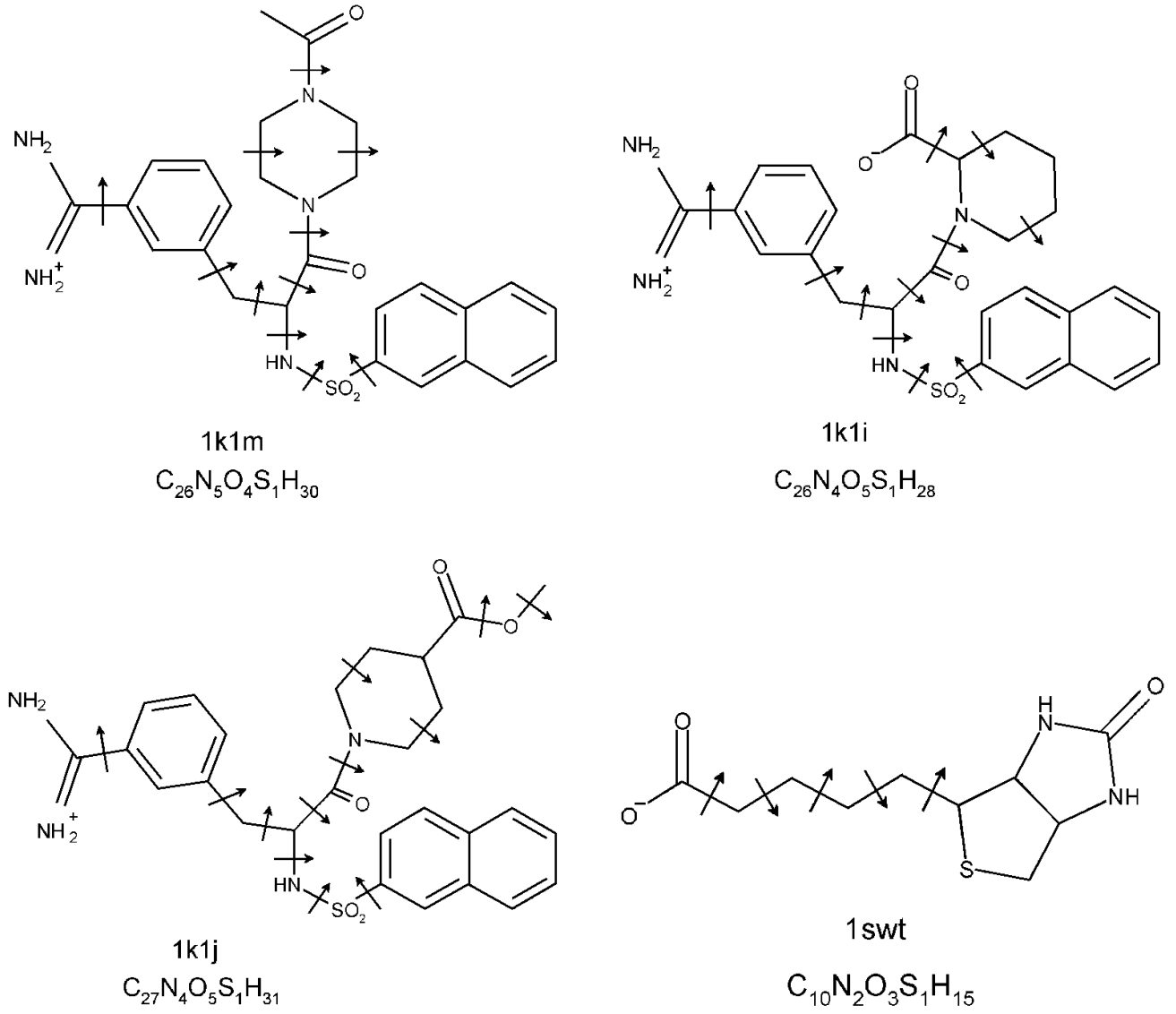

Fig. 2. Structural formulas of ligands for trypsin and streptavidin. Arrows designate the torsion angles where harmonic restraints were applied. Ligand formulas and PDB codes of complexes are provided.

[12-14] for the computation of ΔS_{config}, it is possible to compute by our technique ΔS_{vib} for several tens of conformations. The full value of ΔS_{config} can be calculated as the sum of contributions of ΔS_{vib} with corresponding weights.

To take into account the statistical fluctuations in $T\Delta S_{vib}$, we used the following procedure. The full trajectory was divided in two parts so as the length of each part was sufficient for search in configurational space (300 psec). Value of the $T\Delta S_{vib}$ was computed for each part, and then the arithmetic mean value and root-mean-square error were calculated. This allowed us to estimate possible error of the method at chosen conditions of modeling and processing of the trajectory. Values of the error, which were obtained on this way, were no more than 1 kcal/mol.

The possible dependence of $T\Delta S_{vib}$ on the size of the truncated protein also was investigated. For the trypsin–benzamidine complex, models with 12, 19, and 21 amino acid residues and with the full protein were examined, and even for the smallest model all residues that form hydrogen bonds and specific hydrophobic contacts were retained. Under these conditions, the influence of the size of the protein did not exceed 0.5 kcal/mol. So it is possible to complete the computation of $T\Delta S_{vib}$ using truncated models.

A significant dependence of $T\Delta S_{vib}$ on force constant k_{harm}, which defines the force of harmonic restraints on protein atoms, was revealed. If the value k_{harm} is closed to the value of force constant of bond stretching, the dependence of $T\Delta S_{vib}$ on k_{harm} is weak. However, the reduction of k_{harm} by an order of magnitude or more leads to significant change in $T\Delta S_{vib}$. Also, the value of statistical error of $T\Delta S_{vib}$, which is defined by the scheme

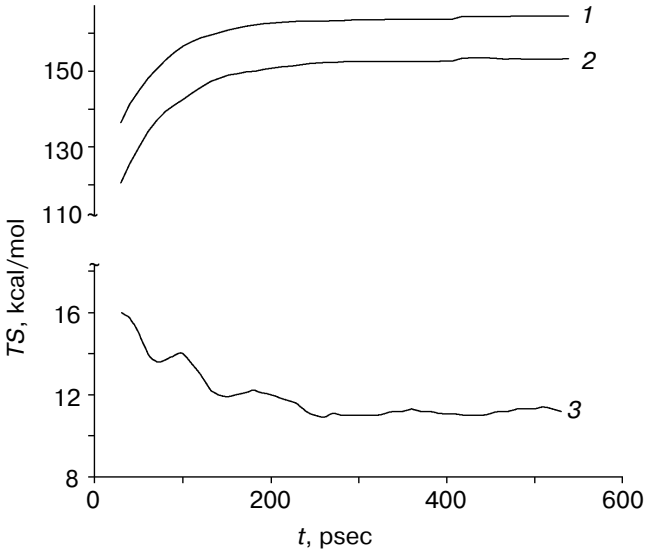

Fig. 3. Dependence of the product of vibrational entropy and temperature on modeling time t for the trypsin–benzamidine complex: *1*) TS_{vib} (PL); *2*) TS_{vib} (P); *3*) $T(S_{vib}(PL) - S_{vib}(P))$.

described above, increases with decrease in k_{harm} and becomes comparable with the value of the vibrational entropy. It seems that this effect is caused by activation of the internal degrees of freedom as a result of the reduction in force of harmonic restraints. To avoid of large fluctuations, we chose values of k_{harm} which are close to values of force constants of bond stretching. A similar investigation was carried out for definition of the dependence of $T\Delta S_{vib}$ on force constant K_φ, which defines the force of harmonic restraints to rotation around bonds for the ligands. It was revealed that starting with K_φ values close to the force constant for valence angle bending, the dependence of $T\Delta S_{vib}$ on K_φ becomes weak.

The results of computations of $T\Delta S_{vib}$ for all nine PL complexes are shown in the table. For PL complexes with benzamidine and its *p*-substituted analogs, we used the truncated structure with 19 amino acid residues, while for large ligands the structure contains more amino acid residues to take into account the formation of hydrogen bonds with Ser195 and Gln192. For the biotin–streptavidin complex, the truncated active site contains 17 amino acid residues. The total number of atoms in all cases was 400-500. For all ligands, $T\Delta S_{vib} > 0$. For benzamidine and its *p*-substituted analogs, the values of $T\Delta S_{vib}$ are significantly more than for ligands with high molecular weight. The possible reason for this is discussed below.

The change in $T\Delta S_{conf}$, calculated using Eq. (4), and also contributions from loss of translational and rotational degrees of freedom $T\Delta S_{tr}$ are shown in the table. Values of $T\Delta S_{tr}$ are close in each group of ligands (benzamidine and its *p*-substituted analogs and large ligands). This is caused by weak logarithmic dependence of $T\Delta S_{tr}$ on mass and inertial moment of the ligand. For small ligands with one or two rotatable bonds, $T\Delta S_{conf}$ is small and does not significantly influence change in entropy. For the large ligands the number of these bonds (arrows on Fig. 2) is great, so $T\Delta S_{conf}$ increases and becomes comparable with other contributions.

The change in configurational entropy $T\Delta S_{config} = T(\Delta S_{tr} + \Delta S_{vib} + \Delta S_{conf})$ is large and negative because $T\Delta S_{vib}$ is too small to compensate for the negative contribution from $T\Delta S_{tr}$ (table). The relatively small values of $T\Delta S_{vib}$ for the four large ligands can be explained as follows. Long length of the ligand leads to large amplitudes of the relative motion of its opposite ends and to increase in the vibrational entropy. For the ligand in the free state there are no restrictions on its motions, while in the bound state amplitudes are restricted by the amino acid residues of the active site. This leads to increase in effective frequencies calculated from CM in the bound state and to reduction of vibrational entropy. For large and flexible ligands, this effect and corresponding decrease in $T\Delta S_{vib}$ will be more significant than for small ones.

Because the harmonic approach is widely used for computation of $T\Delta S_{vib}$ for PL complexes [1, 2, 5, 6], we compared results of the computations of $T\Delta S_{vib}$ by the harmonic approach and by the CM method. All computations were performed in vacuum because the possibility of harmonic analysis with GB is absent in CHARMM. The results are shown in the table.

For PL complexes with benzamidine and its *p*-substituted analogs, in all cases $T\Delta S_{vib.cm}$ exceeded $T\Delta S_{vib.harm}$ by 1-3 kcal/mol, while for the other complexes the reverse tendency was observed: $T\Delta S_{vib.harm}$ exceeds $T\Delta S_{vib.cm}$ by 2-4 kcal/mol. A possible reason for this difference was discussed above and may be caused by the restraints of the amplitudes of vibrations with low frequencies connected with internal degrees of freedom of the ligand. So the harmonic approach can overestimate $T\Delta S_{vib}$ for relatively large ligands and underestimate it for small ligands. Let us note that the vacuum values of $T\Delta S_{vib.cm}$ in all cases are below those obtained using the GB approach. In our opinion, this is caused by reduction in the energy of interaction of charged groups in water in comparison with vacuum, which leads to increase in the amplitudes of ligand vibrations in the active site.

Results of the computations of the entropy change caused by partial desolvation of protein and ligand, which were calculated using Eqs. (5) and (6), are shown in the table. For all the complexes, both $T\Delta S_{pol}$ and $T\Delta S_{nonpol}$ are positive, and the contribution of the nonpolar component exceeds that of the polar component. The main contribution to $T\Delta S_{pol}$ is the volume term of the ligand (Eq. (5)) because the surface area of the PL complexes and the proteins differ slightly.

The change in configurational entropy $T\Delta S_{config}$, the entropic contribution from the effect of partial desolvation of protein and ligand $T\Delta S_{solv}$, their sum $T\Delta S_{calc}$, and also corresponding experimental values are shown in the

Calculated values of the components of the change in the product of temperature and entropy during formation of protein–ligand complexes

Complex	TS_{vib}(P)	TS_{vib}(PL)	TS_{vib}(L)	$T\Delta S_{vib}$	$T\Delta S_{tr}$	$T\Delta S_{conf}$	$T\Delta S_{config}$	$T\Delta S_{vib.cm}$	$T\Delta S_{vib.harm}$
-H	145.6*	156.9	4.9	6.4	–18.6	–0.3	–12.5	3.2	1.4
-OCH_3		160.2	6.7	7.9	–19.0	–0.6	–11.5	1.7	1.0
-CH_3		158.9	5.1	8.2	–19.3	–0.3	–11.7	2.4	0.6
-NH_2		158.4	4.7	8.1	–18.9	–0.3	–11.0	2.2	1.0
-$CONH_2$		160.7	6.4	8.7	–19.5	–0.6	–11.4	4.2	1.2
1k1i	180.8**	222.0	37.8	3.3	–22.5	–3.3	–22.5	–0.9	3.1
1k1j		224.7	43.0	0.9	–22.6	–3.6	–25.3	–1.5	3.5
1k1m		226.1	44.7	1.0	–22.5	–3.3	–24.8	–0.5	3.0
Biotin	176.3***	196.9	20.1	2.5	–21.7	–1.5	–20.7	–1.6	0.5

Complex	ΔG_{pol}	$T\Delta S_{pol}$	A(PL)	A(P)	V(L)	$T\Delta S_{nonpol}$	$T\Delta S_{solv}$	$T\Delta S_{calc}$	$T\Delta S_{exp}$
-H	173	3.5	1814	1821	382	7.8	11.3	–1.2	1.8
-OCH_3	172	3.4	1808		451	9.3	12.7	1.2	0.5
-CH_3	175	3.5	1810		423	8.7	12.2	0.5	2.3
-NH_2	185	3.7	1811		422	8.6	12.3	1.3	2.2
-$CONH_2$	181	3.6	1810		481	9.8	13.4	2.0	2.8
1k1i	211	4.2	2189	2132	1179	18.2	22.4	–0.1	–1.3
1k1j	261	5.2	2182		1218	19.3	24.5	–0.8	1.0
1k1m	234	4.7	2174		1208	19.7	24.4	–0.4	2.0
Biotin	210	4.2	2086	2125	624	14.4	18.6	–2.1	–7.8

Note: Here and below, only R-groups are shown for benzamidine and its *p*-substituted analogs. Truncated structures of the complexes contain, respectively, *19, **22, and ***17 amino acid residues. Values of the product of change in entropy and the temperature and also values of the change in polar term of solvation free energy (ΔG_{pol}) are given in kcal/mol, surface areas A(PL) and A(P) in Å^2, and ligand volume (V) in Å^3.

table. In all cases investigated, negative contribution from $T\Delta S_{config}$ is essentially compensated by the positive value of $T\Delta S_{solv}$. The value of root mean square deviation of calculated and experimental values (with the exception of the biotin–streptavidin complex) is 1.6 kcal/mol. Possible reasons for differences in $T\Delta S_{calc}$ and $T\Delta S_{exp}$ are due to our approach for computation of the vibrational entropy (it was calculated for only the conformation of protein, ligand, and PL complex) and also by the simplified scheme of consideration of entropic contribution from change in polar part of the free energy of solvation.

The most significant difference in $T\Delta S_{calc}$ and $T\Delta S_{exp}$ was found for the biotin–streptavidin complex. The excess of $T\Delta S_{calc}$ over $T\Delta S_{exp}$ is due to the underestimation of the configurational entropy of the streptavidin in the free state. The active site of streptavidin is more flexible than that of trypsin. For instance, the mobility of Tyr43 and Ser88, which is limited in the bound state by hydrogen bonds with biotin, can be more significant in the free state, and this leads to increase in configurational entropy of the protein and decrease in $T\Delta S_{config}$. So in this case using only the conformation in our method can leads to relatively large inconsistency.

Our technique can be implemented to solve the problem of virtual screening of organic compounds with the goal of selection of those which form a PL complex with the target protein (see, for instance, [33]). To carry out virtual screening, it is first necessary to "dock" the ligand to the active site of the target protein. The docking procedure requires sampling of a large number (10^6-10^7) of geometric configurations of the PL complex and computation of the energy for each of them. Configurations selected by this procedure can be used as a starting point for MD modeling. Change in the enthalpy ΔH_{calc} due to PL complex formation can be calculated by means of the technique described in [34], which is based on quantum chemistry modeling of the active site of the protein and the ligand in free and bound states while taking the solvent into account. So in this way the protein–ligand binding free energy can be calculated as $\Delta G = \Delta H - T\Delta S$.

Docking using programming complexes especially elaborated for this goal (http://dock.compbio.ucsf.edu/Overview_of_DOCK/index.htm,http://www.ccdc.cam.ac.uk/products/life_sciences/gold/,http://autodock.scripps.edu/) requires several hours for a ligand of moderate size. Using our technique, most time (also several hours) is needed to obtain the MD trajectory with length ~500 psec for the computation of $T\Delta S_{vib}$ at fixed conformation. The time required for the computation of ΔH_{calc} [34] for one conformation is similar. When several conformations of similar energy are revealed as a result of docking, time of computation both $T\Delta S_{calc}$ and ΔH_{calc} increase proportionally to the number of conformations. In this case, the computation can be done only for conformations that should be specially selected using additional assumptions.

The entropy contribution $T\Delta S_{calc}$ to protein–ligand (PL) binding free energy has been studied for nine PL complexes. Loss of the translational/rotational degrees of freedom (ΔS_{tr}) is calculated in the ideal gas approximation. Change in the vibration entropy (ΔS_{vib}) is calculated using the effective quantum oscillator approach with frequencies derived from the coordinate covariance matrix, so the inharmonic effects are taken into account. The harmonic restraints were applied to atoms of the protein and to torsions of the ligand to fix the initial conformation under MD modeling. Using such restraints leads to significant decrease (from 10-50 nsec [12-14] to 500 psec) of the MD trajectory length necessary to obtained the correct value of $T\Delta S_{vib}$. It was found that $T\Delta S_{vib}$ is greater for the small ligands than for the large ones. One possible reason for this is the restriction of amplitude of the internal vibrations of the ligand in the bound state. The values of $T\Delta S_{vib}$ calculated in the harmonic approach and using the CM method were compared. For the small ligands – benzamidine and its *p*-substituted analogs – the harmonic approach underestimates $T\Delta S_{vib}$, while for the large ones – in contrast, overestimates.

Change in the solvation entropy $T\Delta S_{solv}$ was computed using its decomposition into polar $T\Delta S_{pol}$ and nonpolar $T\Delta S_{nonpol}$ terms. It was found that for all the investigated complexes $T\Delta S_{nonpol} > T\Delta S_{pol} > 0$, and $T\Delta S_{solv}$ provides the main contribution to the compensation of the entropy decrease from the loss of the translational/rotational degrees of freedom.

The value of root mean square deviation of calculated $T\Delta S_{calc}$ and experimental $T\Delta S_{exp}$ values (with exception of biotin–streptavidin complex) is 1.6 kcal/mol. For the biotin–streptavidin complex the excess of $T\Delta S_{calc}$ over $T\Delta S_{exp}$ (by 5.7 kcal/mol) may be caused by underestimation of the configurational entropy of the streptavidin in the free state.

This work was performed with financial support from the Russian Foundation for Basic Research (grant 06-03-33171).

REFERENCES

1. Chang, C. A., Chen, W., and Gilson, M. K. (2007) *Proc. Natl. Acad. Sci. USA*, **104**, 1534-1539.
2. Chang, C. A., and Gilson, M. K. (2003) *J. Comput. Chem.*, **24**, 1987-1998.
3. Woo, H. J., and Roux, B. (2005) *PNAS*, **102**, 6825-6830.
4. Carlsson, J., and Aquist, J. (2006) *Phys. Chem. Chem. Phys.*, **8**, 5385-5395.
5. Tudor, B., and Karplus, M. (1994) *J. Mol. Biol.*, **238**, 405-415.
6. Schwarzl, S. M., Tschopp, T. B., Smith, J. C., and Fischer, S. (2002) *J. Comput. Chem.*, **23**, 1143-1149.
7. Karplus, M., and Kushik, J. N. (1981) *Macromolecules*, **14**, 325-332.
8. Schlitter, J. (1993) *Chem. Phys. Let.*, **215**, 617-621.
9. Andricioaei, I., and Karplus, M. (2001) *J. Chem. Phys.*, **115**, 6289-6292.
10. Schafer, H., Mark, A. E., and Gunsteren, W. F. (2000) *J. Chem. Phys.*, **113**, 7809-7817.
11. Luo, H., and Sharp, K. (2002) *PNAS*, **99**, 10399-10404.
12. Schafer, H., Daura, X., Mark, A. E., and Gunsteren, W. F. (2001) *Proteins: Struct. Funct. Genet.*, **43**, 45-56.
13. Schafer, H., Smith, L. J., Mark, A. E., and Gunsteren, W. F. (2002) *Proteins: Struct. Funct. Genet.*, **46**, 215-224.
14. Hsu, S. D., Peter, C., Gunsteren, W. F., and Bonvin, A. M. J. J. (2005) *Biophys. J.*, **88**, 15-24.
15. Carlsson, J., and Aquist, J. (2005) *J. Phys. Chem. B.*, **109**, 6448-6456.
16. Dixit, S. B., and Chipot, C. (2001) *J. Phys. Chem. A*, **105**, 9795-9799.
17. Dominy, B., and Brooks, C. L. (1999) *J. Phys. Chem.*, **103**, 3765-3773.
18. Brooks, B. R., Bruccoleri, R., Olafson, B., States, D., Swaninathan, S., and Karplus, M. (1983) *J. Comput. Chem.*, **4**, 187-217.
19. Halgren, T. A. (1996) *J. Comput. Chem.*, **17**, 490-519.
20. Landau, L. D., and Lifshiz, E. M. (2002) *Theoretical Physics* [in Russian], Vol. 5, Fizmatlit, Moscow.
21. Morris, G. M., Goodsell, D. S., Halliday, R. S., Huey, R., Hart, W. E., Belew, R. K., and Olson, A. J. (1998) *J. Comp. Chem.*, **19**, 1639-1662.
22. Tomasi, J., and Persico, M. (1994) *Chem. Rev.*, **94**, 2027.
23. Alexandrovsky, V. V., Basilevsky, M. V., Leontyev, I. V., Mazo, M. A., and Sulimov, V. B. (2004) *J. Phys. Chem. B*, **108**, 15830.
24. Basilevsky, M. V., Grigoriev, F. V., Leontyev, I. V., and Sulimov, V. B. (2005) *J. Phys. Chem. A*, **109**, 6939.
25. Huang, D. M., Geissler, P. L., and Chandler, D. (2001) *J. Phys. Chem. B*, **105**, 6704-6709.
26. Grigoriev, I. S., and Meilikhov, E. Z. (eds.) (1991) *Physical Values* [in Russian], Energoatomizdat.
27. Meissner, T., and Wentz, F. J. (2004) *IEEE Trans. Geosci. Remote Sensing*, **42**, 1836-1849.
28. Klotz, I. M. (1997) *Ligand–Receptor Energetics*, Jonh Wiley & Sons Inc., N. Y.
29. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne, P. E. (2000) *Nucleic Acids Res.*, **28**, 235-242.
30. Word, J. M. (1999) *J. Mol. Biol.*, **285**, 1735-1747.
31. Dullwebwr, F., Stubbs, M. T., Musil, D., Sturzebecher, J., and Klebe, G. (2001) *J. Mol. Biol.*, **313**, 593-614.
32. Nikitina, E. A., Sulimov, V. B., Zayets, V., and Zaitseva, N. (2004) *Int. J. Quantum Chem.*, **97**, 747-763.
33. Eckert, H., and Bajorath, J. (2007) *Drug Discovery Today*, **12**, 225-233.
34. Nikitina, E. A., Sulimov, V. B., Grigoriev, F. V., Kondakova, O. A., and Luschekina, S. V. (2006) *Int. J. Quantum Chem.*, **106**, 1943-1963.