Ukrainica Bioorganica Acta 2 (2006) 47—55

Ukrainica Bioorganica Acta

www.bioorganica.org.ua

Application of distribution function of rotation and translation degrees of freedom for CK2 inhibitors K_i estimation

O.Ya. Yakovenko, A.G. Golub, V.G. Bdzhola*, S.M. Yarmoluk

Institute of Molecular Biology and Genetics, NAS of Ukraine
150 Zabolotny Str., Kyiv, 03143, Ukraine

Summary. The method of estimating dissociation constants from distribution functions of rigid body translation and rotation degrees of freedom has been developed. Two ways of K_i estimations have been proposed. The first approach, arbitrary model (AM), is derived directly from statistical physics equations. The results show significant overestimation of binding free energy. The second one, quasy-arbitrary model (QM), was derived from assumption about non-symmetrical shape of bound ligand configuration space. The correct ranking of ligands with the similar binding mode was obtained for common distribution theory equations, but the QM K_i estimation method was found to be more appropriate than AM method. The predictive results of method were tested with a set of mainly rigid CK2 inhibitors with resolved crystal structures.

Keywords: distribution function, configuration space, rotation-translation, freedom degree, free energy, K_i estimation.

Introduction. Free energy of binding is the cornerstone of molecular association/dissociation description. The main aim of all up-to-date virtual screening techniques and the most of molecular dynamic analysis is an estimation of free energy of binding. A lot of strategies were proposed to perform such estimation perfectly. So far the common problem of all known strategies is that sufficiently accurate methods require too many intensive computations to be used in high throughput virtual screening routines.

Different approaches that vary in sophistication and complexity have been used to calculate free energies of binding in «receptor-ligand» system. Virtual screening of large compound databases usually requires simplified scoring schemes to avoid huge time expenses [15]. The free energy of binding may be estimated on the basis of a continuum solvent approximation assuming quadratic fluctuations around a unique configuration [6]. Basic components may include steric energy and electrostatic energies, sometimes supplemented by other terms accounting for hydrogen binding and solvation effects [3, 4, 9, 13, 16].

A popular approach is the Molecular Mechanics Poisson-Boltzmann (PB) and Surface Area (MM_PB-SA) method that relies on a mixed scheme combining configurations sampled from molecular dynamics (MD) simulations with explicit solvent, together with free energy estimators based on an implicit continuum solvent model [10]. Similarly, the Linear Interaction Energy method also uses averages calculated from explicit solvent simulations within a linear response framework [1]. Despite the usefulness of these methods, they do not offer a rigorous route to compute the equilibrium binding constant.

The partitioning of binding affinity into several additive terms or descriptors is a widely accepted assumption for the development of empirical regression-based scoring functions.

***Corresponding author.**
Tel./fax: +38044-5222458
E-mail address: hive@ukr.net

Usually a number of empirically derived contributions is fitted to a data set of experimental observations [5, 11, 14]. Approaches such as VALIDATE are based on the ideas of QSAR. Precision of K_i prediction with these approaches is about 1.5 orders of magnitude.

With respect to the precise calculation in K_i prediction of relative binding energies, the free energy perturbation (FEP) [8] is currently the only method that attempts to deal seriously with calculating ensemble averages and considers the behavior of solvent molecules explicitly. However, despite various approximations directed towards performance enhancement, this method is computationally intensive and restricted to small molecular systems.

The fastest way for dissociation constant calculation is distribution function from statistical physics. The nice accuracy of distribution function in provided by a method described in [2, 12].

Current work is devoted to the development of receptor oriented virtual screening approach based on distribution function. Since it was aimed at the drug design application, the main emphasis was made to choose the fastest approach for individual compound evaluation. To satisfy these requirements the approach with analytical estimations of distribution integrals was chosen. For the validation of this approach we selected the set of known mainly rigid CK2 inhibitors with resolved crystal structures. All these inhibitors are ATP competitive and they are non-covalently bound to ATP binding site. The value of K_i was chosen as the most adequate parameter for predicted and experimental data consistency tests. K_i calculations have been carried out in two ways. The first way was obtained via integration potential energy over all degrees of freedom. The second one was derived as assumption about strongly determined path of ligands dissociation. It supposes a key role of steric restrictions of proteins binding site that ligands can overcome along only one trajectory.

The same ranking of K_i values obtained by both approaches was observed and it was well correlated with the experimental data. Moreover, each of them was revealed to identify crystal position from a random configuration set with satisfactory accuracy.

Theory. According to the well known distribution law of statistical mechanics, Z function of molecular distribution is based on additive integrals of mechanical description of the molecular system. This immediately follows from the logarithmic additive of various states of probabilities and very limited number of constant additive integrals in mechanics: impulses, their momentums, and energies. The macro-property which could be evaluated from such distribution law is the affinity estimation from 3D coordinates of the resolved structure or form the set of modeled 3D complexes. The free energy of molecular binding can be expressed in terms of distribution function as

$$\Delta G = RT \ln \frac{Z_b}{Z_f} \qquad (1),$$

where ΔG — Gibbs free energy of binding, R — gas constant, T — temperature, Z_b — probability of the particular molecule to be in the bound state, Z_f — probability of this molecule to be in any other state (free state). For known 3D complexes we suppose the presence of the molecules in two states only — bound to the receptor in resolved way and unbound state. For modeled 3D complexes we have a set of different configurations. Free energy is the property of the whole set of sophisticated complexes which are a part of configuration space of molecules — complex members. Modified equation here is

$$\Delta G = RT \ln \frac{\sum Z_b}{Z_f} \qquad (2),$$

where we employed the simple rule of probabilities summation in logarithmic manner over several bound states.

At the same time free energy of binding can be easily expressed in terms of dissociation constant as

$$\Delta G = RT \ln K \qquad (3),$$

where K is dissociation constant of ligand-receptor complex. Replacing ΔG in Eq. 3 with equation Eq. 1 results in dissociation constant expression

$$K = \frac{Z_b}{Z_f} \qquad (4).$$

According to statistical physics we can write the distribution function in terms of additive integrals combination of n degrees of freedom:

$$Z = \frac{1}{h^n} \int e^{-\beta(T(p)+U(q))} dpdq \qquad (5),$$

where h — quantum unclearity, β — inverse

temperature factor $\beta = \frac{1}{kT}$, p — impulses, q — coordinates, $T(p)$ — kinetic energy component, $U(q)$ — potential energy component. Equation (Eq. 5) can be integrated by kinetic part and rewritten as

$$Z = \sqrt{\frac{(2\pi)^6 \prod_{i=1}^{6} I_i}{h^{12}\beta^6}} \int e^{-\beta U(q)} dq \qquad (6),$$

where I — impulse moments. Hereinafter, we assume that there should be only 6 essential freedom degrees for known CK2 ligands, namely, three rotation and three translation ones and no internal degrees of freedom allowed. The potential energy part can be integrated from decomposition of potential energy function in Taylor series up to the third degree term. Bilinear form $A = q^T A_{ij} q$ forms such a decomposition that represents the potential energy function around local minima, where A_{ij} is the second derivative of energy over q_i and q_j coordinates. Matrix A can be diagonalized then into diagonal matrix B with orthogonal transformations. Here we used singular value decomposition of A to obtain B. It is a known analytical integral of type $\int_{-\infty}^{+\infty} e^{-k*x^2} dx = \sqrt{\frac{\pi}{k}}$, which is used for analytical solution of equation (Eq. 4). Thus, we obtain equation for Z function evaluation:

$$Z_b = e^{-\beta U_b} \left(\sqrt{\frac{(2\pi)^6 \prod_{i=1}^{6} I_i}{h^{12}\beta^6}} \right) \left(\sqrt{\frac{(2\pi)^6}{\beta^6 \prod_{i=1}^{6} b_{ii}}} \right) \qquad (7),$$

where U_b — the first term of Taylor series of energy function decomposition for bound case.

Product of all self values of matrix A is just a determinant of A (let us denote DetA as D). Usually it can be evaluated in more efficient way (LU decomposition for example) than via matrix diagonalization (for example Housholder method). The values of b, however, can be used in other analyses.

For the reference of unbound state Z function can be approximated with zero U matrix of free energy derivatives and

$$Z_f = e^{-\beta U_f} \sqrt{\frac{(2\pi)^6 \prod_{i=1}^{6} I_i}{h^{12}\beta^6}} \left(8\pi^3 V\right) \qquad (8),$$

where V is a space volume per one molecule in unbound state that is useful to be evaluated in mol units as

$$V = \frac{10^{-3}}{N_A} = 1.66 \cdot 10^{-27} \left(m^3\right) = 1660 \text{ Å}^3$$

U_b from Eq. 7 can be decomposed as

$$U_b = U_L + U_R + U_{LR} + U_{sol} \qquad (9),$$

where U_L — internal ligand energy, U_R — internal receptor energy, U_{LR} — interaction energy between ligand and receptor, U_{sol} — desolvation free energy, which arises as an addition work during ligand transferring from solvent into receptor cavity. U_f therefore decomposes into sum of U_L and U_R only. Currently we assume U_{sol} to be zero, however, we are going to account it in future. According with that one can replace U_b with U_{LR} but we left U_b to keep the theory comprehensive.

So the final formulas for affinity measurement are

$$K = \frac{Z_b}{Z_f} = \exp\left(3\ln\frac{RT}{l} - 0.5\ln D + \frac{U_b}{RT}\right) \qquad (10),$$

where l is a linear length of volume box corresponding to 1660 A^3. This equation corresponds to the canonical statistical physics distribution that we will call the arbitrary model of complex configuration space volume in rotation-translation approximation (AM).

Real calculations of Eq. 10 show a strong disagreement with experimental data (nearly the second power, approximately the longest normal mode component). That is why quasi one-dimensional model (QO) was proposed in contrast to arbitrary multi-dimensional model (AM) expressed in Eq. 10. The basic idea of QO is that only one dissociation coordinate exists — reaction coordinate. Thus, in the B matrix we have to account only a half of possible directions for distribution function evaluation — the second half is reduced by fast energy growing when the ligand tends toward protein cavity walls. So determinant of the matrix B becomes:

$$D` = \sqrt{D} \qquad (11),$$

and dissociation constant in QO model is:

$$K_q = \frac{Z_{b_q}}{Z_f} = \exp\left(3\ln\frac{RT}{l} - 0.25\ln D + \frac{U_b}{RT}\right) \qquad (12).$$

U function from MMFF94 force field has been used in our model. This field has kcal scale of energy evaluation and multiplier 1 kcal = 4184 J should be used to deal with SI system. This is a pair of additive, distance dependent functions so all derivatives in derivative matrix A could be calculated via common equation.

Applyng U function from MMFF94 force field as:

$$U_{ij} = U_{wdv} + U_{elec} = \varepsilon_{ij}\left(\frac{1.07R_{ij}}{r+0.07R_{ij}}\right)^7\left(\frac{1.12R_{ij}^7}{r^7+0.12R_{ij}^7}-2\right)+\frac{332.071q_iq_j}{D(r+\delta)} \quad (13),$$

where e_{ij} and R_{ij} are type dependent atom constants, q_i and q_j are atom charges and r is the distance between atoms, the first and second derivatives of energy function over distance are

$$\frac{dU}{dr} = -\frac{332.071q_iq_j}{D(r+\delta)^2} - \frac{12.5893\varepsilon_{ij}R_{ij}^{14}r^6}{\left(r+0.07R_{ij}\right)^7\left(r^7+0.12R^7\right)^2}\cdot$$
$$-\frac{11.2405\varepsilon_{ij}R_{ij}^7\left(\frac{1.12R_{ij}^7}{r^7+0.12R_{ij}^7}-2\right)}{\left(r+0.07R_{ij}\right)^8} \quad (14)$$

and

$$\frac{d^2U}{dr^2} = \frac{664.142q_iq_j}{D(r+\delta)^3} + \frac{\varepsilon_{ij}}{\left(r+0.07R_{ij}\right)^9}\left(\frac{176.251R_{ij}^{14}r^6\left(r+0.07R_{ij}\right)^2}{\left(r^7+0.12R_{ij}^7\right)^2}\right.$$
$$\left(\frac{r^6}{\left(r^7+0.12R_{ij}^7\right)}+\frac{1}{\left(r+0.07R_{ij}\right)}-\frac{0.42857}{r}\right)+89.9238R_{ij}^7$$
$$\left.\left(\frac{1.12R_{ij}^7}{r^7+0.12R_{ij}^7}-2\right)\right) \quad (15),$$

respectively.

The matrix of second derivatives is calculated as:

$$D_{ij} = \frac{1}{r^2}\left(\frac{d^2U}{dr^2}-\frac{1}{r}\frac{dU}{dr}\right)\frac{\partial r^2}{2\partial_j}\frac{\partial r^2}{2\partial_i}+\frac{1}{r}\frac{dU}{dr}\frac{\partial^2 r^2}{2\partial_j\partial_i} \quad (16),$$

where r is a distance between points, i and j runs over all freedom degrees. The square function of r is chosen to simplify evaluation of the partial derivatives.

Results and discussion. We have developed and successfully tested method of dissociation constant prediction from transition-rotation distribution functions using the set of known CK2 inhibitor complexes. By the time of investigation there were only 10 complexes available in protein structure data bank (1zog, 1zoh, 1zoe, 1om1, 1m2r, 1m2p, 1m2q, 1j91, 1foq, 1lp4). As we were interested in drug design application of our method of dissociation constant calculation, we skipped the evaluation of 1zoh, 1zoge and 1lp4. 1zoh and 1zoge complexes have both Na^+ Cl^- ion pair and small organic molecular inhibitor in the active site. There is a not enough accurate technique in the receptor oriented high throughput virtual screening that would allow modeling ions binding during small organic ligand binding. As we did not assume the analysis of such multicomponent complexes in the receptor based virtual screening, we rejected these structures from the validation set. The ATP complex 1lp4 is the some occurence because of Mg^{2+} constitutive ion presence. Moreover, ATP molecule has a huge amount of rotable bonds that can hardly be evaluated with satisfactory results by current method that does not take into account any internal molecular degrees of freedom.

Speaking about adequacy criteria, we should pay attention to the nature of today CK2 inhibitors. Usually they are rigid molecules with a very restricted amount of rotable bonds in structure. Moreover, 1j91 has no such degrees at all, and all besides 1om1 and 1zog have proton participated rotable bonds only, which are not assigned as rotable bonds by the most of the rotable bonds definition techniques due to miserable mass relation of proton to other parts of compounds and well known protons mobility. Similarly in 1m2p, 1m2q, 1m2r the symmetric nitro or amino groups are present only. That is why we can neglect internal degrees of freedom for nearly all compounds in set. On the other hand, six dimension rotation and translation components of free energy second derivatives matrix describe ligands as a rigid-body system without any additional internal degrees of freedom. From a mechanical point of view, we should account as many degrees of freedom as there are valence bonds and angles in compound, but we can omit account of valence bonds and angles vibrations because of their incomparably high vibration frequency. The torsion dihedral degrees of freedom cannot be omitted because of their vibration period that is close to the time of binding energy barrier overcoming.

Observation of sets of rotable bonds amount is presented in Table 1. This parameter can be used as a scale of methods error. The first numbers in rotable bond column of the table are real numbers of dihedral torsions angles degrees of freedom and the second ones are the numbers of significant rotable bonds in compound. It is easy to see that only two compounds of the set have an insufficient description in rotation-translation terms of distribution functions and significant prediction error should be estimated for them.

The other significant limitation of current model is the assumption that interaction energy

Table 1

K_i estimations for CK2 inhibitors set

PDB Name	ΔU, kcalxmol^{-1}	-0.5 lnD	Lg (K_i), predicted	K_i (OG model), 10^{-6}	Rotable bonds	K_i, zea maise, 10^{-6}
1f0q	-10.6637	-18.0201	-11.7347	1.3573	4/0	1.85
1j91	-12.8915	-15.7910	-10.7666	4.1372	0/0	0.4
1m2p	-17.7921	-17.3588	-11.4475	1.8891	3/0	0.78
1m2q	-6.5671	-17.2513	-11.4008	1.9935	3/0	0.8
1m2r	-10.8699	-18.0800	-11.7607	1.3172	4/0	0.42
1om1	-6.2530	-19.1319	-12.2175	0.7785	3/2	0.17
1zog	-11.6883	-17.6606	-11.5786	1.6245	2/1	0.07

* *q — data concerning QO model.*

of the compound in unbound state is zero. As we did not want to mix results of current method with other methods of molecules affinity prediction, we have chosen a vacuum as a model of unbound state of the molecule. Currently we just admit a strong hydrophobic potential of two structures in the set 1j91 and 1zog. Such structures were found to be underestimated with distribution functions in rotation-translation approximations with neglecting of interaction energy.

The force field used in this work was YFF. It is a slightly modified MMFF94 force field [18] adopted for effective application in drug design by integration together with a charge definition scheme [20]. To integrate YFF we have reformulated atom typification to optimize fit for both electrostatic and geometrical properties. The atom type number is reduced in YFF to 70 and the atom types are defined on the base of electron hybridization state. However, there are 10 types of hydrogen atoms for obtaining better properties fit because hydrogen is usually the most frequent atom in organic compounds. The atom type compiler in YFF is configured to use the descriptors of electronic state instead of topological descriptors as defined in MMFF94. This opportunity permits to use YFF as naturally integrated sophisticated charge definition scheme by Oliferenko with MMFF94 interaction energy functions and basic force constants parameterization by Halgren. The MMFF94 force field was chosen as the most appropriate basic model of molecule interaction as the force field that was parameterized against the resolved crystal structures of small molecules, their interaction and conformation energy.

Distribution function evaluations have to be performed at the minimum extreme point of system energy function defined by Eq. 13. Eq. 6 and Eq. 7 establish this requirement. It means that the minimization of scalar function of vector argument should be performed up to very small gradient length. The gradient of energy function defined by Eq. 13 is a sum over all atom pairs of the protein-ligand complex of function Eq. 14.

The systems minimization was performed in two stages. The first stage was carried out with our implementation of the steepest descent and l-bfgs [17] algorithms in Cartesian space. Such relaxation technique was applied to the whole system of protein, ligand and co-crystallized water molecules. After global minimization had converged to linear search step length of 10^{-8}A, ligand position optimization in rotation-translation space was performed.

As we need six variables only to define relative ligand position in 3D space, the method of Newton was chosen as the most appropriate. Newton method of system minimization at iterative step k is defined as

$$x^{k+1} = x^k - d^2 f^{-1} df \qquad (17).$$

Three angle variables of the rotations around three coordinate axes were chosen. Such choice has a preference in explicit evaluation of each first derivative of axis associated rotation freedom degree, when Euler angles formalism, for example, omits a direct evaluation of derivative of rotation around applicator axis. In such angles system rotational matrix is defined as a product of three sequential rotations:

$$R = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\alpha & \sin\alpha \\ 0 & -\sin\alpha & \cos\alpha \end{bmatrix} \begin{bmatrix} \cos\beta & 0 & -\sin\beta \\ 0 & 1 & 0 \\ \sin\beta & 0 & \cos\beta \end{bmatrix} \begin{bmatrix} \cos\gamma & \sin\gamma & 0 \\ -\sin\gamma & \cos\gamma & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

and in simplified form

$$R = \begin{bmatrix} \cos\beta\cos\gamma & \cos\beta\sin\gamma & -\sin\beta \\ \sin\alpha\sin\beta\cos\gamma - \cos\alpha\sin\gamma & \cos\alpha\cos\gamma + \sin\alpha\sin\beta\sin\gamma & \sin\alpha\cos\beta \\ \cos\alpha\sin\beta\cos\gamma + \sin\alpha\sin\gamma & -\sin\alpha\cos\gamma + \cos\alpha\sin\beta\sin\gamma & \cos\alpha\cos\beta \end{bmatrix}$$

(18).

So position derivatives of ligand atoms are defined as $\frac{\partial r}{\partial \alpha} = \frac{\partial r^2}{2r\partial \alpha}$ and

$$r^2 = \vec{r}_{ij} \cdot \vec{r}_{ij} = \left(R\vec{r}_i - \vec{r}_j\right)^T \cdot \left(R\vec{r}_i - \vec{r}_j\right) =$$

$$\vec{r}_i^{\,T} R^T R\vec{r}_i - \vec{r}_j^{\,T} R\vec{r}_i - \vec{r}_i^{\,T} R^T \vec{r}_j + \vec{r}_j^{\,T}\vec{r}_j \qquad (19).$$

The R is a skew-symmetric matrix, so the angles derivatives can be rewritten in vector form as

$$\frac{\partial r^2}{\partial \alpha} = -\vec{r}_j^{\,T}\left(\frac{\partial R}{\partial \alpha}\right)\vec{r}_i - \vec{r}_i^{\,T}\left(\frac{\partial R}{\partial \alpha}\right)^T \vec{r}_j = -2\vec{r}_j^{\,T}\left(\frac{\partial R}{\partial \alpha}\right)\vec{r}_i \quad (20).$$

The linear translation derivatives are simply defined as

$$\frac{dr}{dx} = \frac{x}{r} \qquad (21).$$

The second derivatives of distance between points are defined over two translation degrees

$$\frac{\partial^2 r^2}{\partial x_i \partial x_j} = \delta_j^i \qquad (22),$$

where δ is a *Kronecker* symbol, over two rotation degrees as

$$\left(\frac{\partial^2 r^2}{\partial\alpha\partial\alpha}\right) = -2\vec{r}_j^{\,T}\left(\frac{\partial^2 R}{\partial\alpha^2}\right)\vec{r}_i \qquad (23),$$

mixed second derivatives by rotation and translation degrees

$$\frac{\partial^2 r^2}{\partial x\partial\alpha} = -2\vec{r}_j^{\,T}\left(\frac{\partial R}{\partial\alpha}\right)\left(\frac{\partial \vec{r}_i}{\partial x}\right) \qquad (24).$$

Eq. 21-25 together with Eq. 14, 15 can be substituted with Eq. 16 and form the second order derivative matrices required for Newton method of function minimization Eq. 17. Minimization in rotation-translation coordinates performed with this approach resulted in extremely deep minimization up to gradient of 10^{-4} kcallxmol^{-1} A^{-1} of ligand interaction energy function. This fact eliminates any doubts about the quality of extreme point location for calculations of distribution function.

Evaluation of distribution functions was performed for a model of three independent translation and rotational degrees. Rotation degrees were chosen similarly to minimization approach to handle the rotations around each axis. Similar equations to Eq. 20-24 were used except rotation matrix of Eq. 18 that was substituted with matrices (or their pairs in case of second derivatives) of elementary rotations. It is easy to see that the second derivatives matrix in such case will be symmetrical instead of asymmetrical Eq. 18. All the energy functions used were the same as in rotation-translation minimization routine. Interaction energy of ligand atoms with all other atoms of molecular system was evaluated in extreme point as well.

The results are shown in Table 1. Two columns (third and fourth) illustrate the obtained values of interaction energies and distribution functions correspondingly. The interaction energy was calculated as a sum of atom pairs interaction of all ligand atoms with all receptor atoms located within 10-12A around ligand. In the YFF the elementary interaction unit is not an atom but the atomic group without any internal freedom degrees in it — an anchor. So all protein anchors that have at least one atom at the distance below 10A from ligand were summated into ligand protein interaction energy. This sum is printed in the third column of Table 1. Each atom pair interaction energy was calculated with Eq. 13. The fourth column is formed by calculating the determinate of the second derivatives energy matrix formed as described above. Since the only difference of the distribution functions of different ligands results from different values of their configuration space, the only item of -0.5l nD is different (where D is a determinant of second derivatives matrix). This data is stored in the fourth column of Table 1. As any configuration space relations this one is measured not only in item scale but in energy units (kcall mol^{-1}) as well.

As mentioned above, the MMFF94 field is one of the best fields for receptor-drug interaction calculations. However, in our experiments the interaction energy calculated with this force field shows certain mismatch with experimental data. The YFF simplification can not be the origin of such mismatches because it uses the same nonbond potential parameters of atoms and the same energy functions of MMFF94. However, the desolvation energy is suggested to be a significant correction of final energy results. Nevertheless, it is known that force fields have significant problems with estimation of interaction energy and some aditional correction functions are necessary to obtain a fit with experimental data [19]. Our calculation results suggest the same problem in MMFF94 with evaluating

of protein-ligand interactions. To fit above suggestions with experimental data of CK2 inhibitors activity we divided the set into two groups. The first group was formed of compounds that have strong electrostatic interactions in their binding mode. In our set such compounds are 1om1, 1m2r, 1m2p and 1m2q. The second group of compounds was formed of 1zog and 1j91 that bind to protein via strong hydrophobic potential. These compounds fail in both energy and distribution calculations. Moreover, they have a significantly different binding mode — they are the only compounds in the set that have various hydrophilic groups located inside protein hydrophobic cavity without any constitutive water molecule. As for distribution function, we have to admit that all compounds but 1f0q have a certain direction to leave the site. 1f0q is totally surrounded with protein — solved in protein — and that is why we suppose the distribution function of protein as the main entropy change during binding. However, we can not evaluate flexible protein distribution functions within rotation-translation formalism.

The lack of agreement of values of YFF energy functions with experimental data made us remove interaction energy term from distribution function equation. As a result we lost the opportunity for evaluation of relations between the different binding modes completely. However, the manual grouping of ligands into binding mode permits us to assume the interaction energy to be the same for similar ligands within the same binding mode. This gives us the possibility to evaluate relative ligand-ligand K_i with assumption for their similar interaction energy values. Such «tuning» of ligands activity is especially interesting for drug optimization applications when the single atom mutation can shift compounds K_i in order of magnitude scale. Usually such mutations do not have significant changes in interaction energy and can hardly be evaluated without applying time-consuming free energy perturbation techniques.

Unfortunately, it was found that K_i values calculated with Eq. 10 (column 5 of Table 1) show nearly the second power overestimation of experimentally derived K_i values. The possible explanation of this fact is a structure of protein as a complex 3D object. This means that left and right energy parabolas branches are not symmetrical and there are entrance side and dead end side of ligands motions in receptors cavity for each freedom degree. There are two conclusions from this fact. The first is that the correct estimation of ligands distributions functions is expected from odd series of energy function approximations whose integral over infinite can not be evaluated analytically (it tends to infinity). The second conclusion is illustrated in our approach of Eq. 12. We suggest significant reduction of the dead end branch of configuration space parabola due to exponential growing of repulsive term of van-der-waals interaction energy. So we have to reduce the configuration space volume almost twice. The K_i calculated with this equation is shown in the six column of Table 1. As expected, they have a bit lower values than experimental K_i due to the absence of interaction energy term that was not correlated (see column 3 of Table 1).

Both forms of K_i evaluation show correct ranking of the ligands of the same binding mode. The cross-relations of compounds in predicted K_i are shown in Table 2, which consists of two parts — the first part is relations of different CK2 inhibitors in their experimentally derived activities. The second part of Table consists of predicted relations calculated from distribution function. As distribution function values are in logarithmic scale, this part of Table was formed from the fourth column of Table 1 with exponentiation of the corresponding values difference. However, MMFF94 interaction energy values do not show even correct ranking both inside and between groups. So we do not include it into our Table.

The distribution function shows correct ranking for both groups of ligands. The relations with error lower than two times the experimental K_i are highlighted in bold. These values are grouped into hydrophilic and hydrophobic binding mode. The former is a pair of 1j91 and 1zog, the latter is 1m2r, 1m2p, 1m2q and 1om1. The distribution function is supported by a more sophisticated approach than single energy function evaluation. It involves the differences formalism and depend on Hamiltonian formulation less than simple energy function evaluation can

Table 2

The relation of experimentally derived ligands K_i with calculated by distribution function

		0.4	**0.07**	**0.78**	**0.8**	**0.42**	**0.17**	**1.85**
1j91	**0.4**	1	0.175	1.95	2	1.05	0.425	4.625
1zog	**0.07**	**5.714286**	1	11.14286	11.42857	6	2.428571	26.42857
1m2p	**0.78**	0.512821	0.089744	1	**1.025641**	**0.538462**	**0.217949**	2.371795
1m2q	**0.8**	0.5	0.0875	**0.975**	1	**0.525**	**0.2125**	2.3125
1m2r	**0.42**	0.952381	0.166667	**1.857143**	**1.904762**	1	**0.404762**	4.404762
1om1	**0.17**	2.352941	0.411765	**4.588235**	**4.705882**	**2.470588**	1	10.88235
1f0q	**1.85**	0.216216	0.037838	0.421622	0.432432	0.227027	0.091892	1
		-15.791	**-17.661**	**-17.359**	**-17.251**	**-18.08**	**-19.132**	**-18.02**
1j91	**-15.791**	1	**0.154185**	0.208503	0.232167	0.101368	0.035405	0.107625
1zog	**-17.661**	**6.485702**	1	1.352291	1.505763	0.657441	0.229627	0.698025
1m2p	**-17.359**	4.796085	0.739486	1	**1.113491**	**0.486169**	**0.169806**	0.51618
1m2q	**-17.251**	4.307252	0.664115	**0.898077**	1	**0.436617**	**0.152499**	0.463569
1m2r	**-18.08**	9.865068	1.521049	**2.0569**	**2.290339**	1	**0.349273**	1.06173
1om1	**-19.132**	28.24454	4.354893	**5.889081**	**6.557438**	**2.863086**	1	3.039825
1f0q	**-18.02**	9.2915	1.432613	1.937309	2.157176	0.941859	0.328966	1

** ligand relations that do not correlate with each other are presented in regular; ** ligand relations that satisfy experimental data are presented in bold.*

be increased with any constant without losing physical meaning.

Conclusions. The method of rigid body rotation-translation distribution function has been developed and successfully tested on the set of resolved CK2 inhibitors. The YFF modification of MMFF94 was applied. The ultimate rotation-translation minimization strategy with Newton method allows to identifing the minimum point with gradient below 10^{-4} kcalxmol^{-1} A^{-1}. The distribution function showed correct compounds ranking inside compounds groups with the same binding mode. The MMFF94 energy evaluation function was found to be unacceptable for direct estimation of ligands binding constant without additional scaling. The distribution function approach was found suitable for K_i estimation in form of the half of configuration space values due to complicated 3D structure of proteins active site.

Abbreviations. AM — arbitrary model of ligand-receptor complex configuration space volume in rotation-translation approximation.

QM — non-symmetrical model of ligand-receptor complex configuration space volume in rotation-translation approximation.

Надійшла до редакції 15.12.2006 р.

Застосування функцій розподілу обертально-трансляційних ступенів свободи для передбачення K_i інгібіторів CK2

О.Я. Яковенко, А.Г. Голуб, В.Г. Бджола, С.М. Ярмолюк

Інститут молекулярної біології і генетики НАН України
вул. Академіка Заболотного, 150, Київ, 03143, Україна

Резюме. Розроблено метод для передбачення констант дисоціацій за допомогою функцій розподілу обертальних і трансляційних ступенів свободи. Запропоновано дві моделі передбачення K_i: 1) довільну, отриману прямим розв'язком рівнянь статистичної фізики (результати, здобуті за допомогою цієї моделі, демонструють значну переоцінку при передбаченні вільної енергії зв'язування), та 2) квазі-довільну, виведену згідно з припущенням про несиметричність форми конфігураційного простору ліганду у зв'язаному стані. Правильне сортування лігандів у межах однакових типів зв'язування одержано для спільної частини математичного апарату, але передбачення K_i за допомогою квазі-довільного підходу виявилось більш ефективним. Передбачувальні здатності методу перевірено на вибірці рентген-кристалічних структур комплексів CK2 та лігандів, які переважно не містять обертальних зв'язків.

Ключові слова: функція розподілу, конфігураційний простір, обертально-трансляційний ступінь свободи, вільна енергія, передбачення K_i.

References

1. *Brandsdal B.O., Osterberg F., Almlof M., Feierberg I., Luzhkov V.B. & Aqvist J.* Free energy calculations and ligand binding // Adv. Protein Chem. — 2003. — 66. — P. 123-158.

2. *Benzhuo Lu Chung F. Wong.* (1997) Direct estimation of entropy loss due to reduced translational and rotational motions upon molecular binding // Biopolymers. — 2005. — Vol. 79. — P. 277-285.

3. *Grootenhuis P.D.J., Galen P.J.M.V.* Correlation of binding affinities with non-bonded interaction energies of thrombin-inhibitor complexes // Acta. Cryst. — 1995. — D51. — P. 560-566.

4. *Holloway M.K., Wei, J.M.* A Priori prediction of activity for hiv-1 protease inhibitors employing energy minimization in the active site // J. Med. Chem. — 1995. — 38. — P. 305-317.

5. *Jain A.N.* Scoring noncovalent protein-ligand interactions: A continuous differentiable function tuned to compute binding affinities // J. Comput.-Aided Mol. Design. — 1996. — 10. — P. 427.

6. *Luo H., Sharp K.* On the calculation of absolute macromolecular binding free energies // Proc. Natl. Acad. Sci. USA. — 2002. — 99. — P. 10399-10404.

7. *Roux B., Simonson T.* Implicit solvent models // Biophys. Chem. — 1999. — 78. — P. 1-20.

8. *Kollman P.A.* Free energy calculations: applications to chemical and biochemical phenomena // P. Chem. Rev. — 1993. — 93. — P. 2395-2417.

9. *Luty B.A., Wasserman Z.R., Stouten P.F.W., Hodge C.N., Zacharias M., McCammon J.A.* A molecular mechanics/grid method for evaluation of ligand-receptor interactions // J. Comp. Chem. — 1995. — 16. — P. 454.

10. *Massova I. and Kollman P.A.* Combined molecular mechanical and continuum solvent approach (MM-PBSA/GBSA) to predict ligand binding // Perspect. Drug Discov. Design. — 2002. — 18. — P. 113-135.

11. *Murray C.W., Auton T.R. and Eldridge M.D.* Empirical scoring functions. II. The testing of an empirical scoring function for the prediction of ligand-receptor binding affinities and the use of Bayesian regression to improve the quality of the model // J. Comput.-Aided Mol. Design. — 12. — P. 503.

12. *Nir Ben-Tal, Barry Honig, Carey Bagdassiarian, Avoinoam Ben-Shaul.* Association entropy in adsorption processes // Biophisical Jornal. — 2000. — Vol. 79.

13. *Ortiz A.R., Pisabarro M.T., Gago F.* Prediction of drug binding affinities by comparative binding energy analysis // J. Med. Chem. — 1995. — 38. — P. 2681-2691.

14. *Rose P.W.* Scoring methods in ligand design, Proceedings of 2nd UCSF Course in Computer-Aided Molecular Design, San Francisco, CA, 1997.

15. *Schneider G. & Bohm H.J.* Virtual screening and fast automated docking methods // Drug Discovery Today. — 2002. — 7. — P. 64-70.

16. *Sansom C.E., Wu J., Weber I.T.* Molecular mechanics analysis of inhibitor binding to HIV-1 protease // Protein Eng. — 1992. — 5. — P. 659-667.

17. *Bird R.H., Nocedal J. and Schnabel B.* Representation of quasy-newton matrices and their use in limited memory methods, NAM-03 technical report January 1996.

18. *Halgren T.A.* Merc molecular force field // J. Comp. Chem. — 1996. — Vol. 17. — P. 490-519.

19. *Almlof M., Brandsdal B.O. and Aqvist J.* Binding affinity prediction with different force fields: examination of the linear interaction energy method // J. Comp. Chem. — 2004. — Vol. 25, No 10. — P. 1242-1254.

20. *Oliferenko A.A., Krylenko P.V., Palyulin V.A. and Zefirov N.S.* A new scheme for electronegativity equalization as a source of electronic descriptors: application to chemical reactivity // SAR and QSAR in Environmental Research. — 2002. — Vol. 13, No 2. — P. 297-305.