

## The new method of distribution integrals evaluations for high throughput virtual screening

O.Ya. Yakovenko\*, A.A. Oliferenko<sup>1</sup>, A.G. Golub,  
V.G. Bdzhola, S.M. Yarmoluk

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

<sup>1</sup>*Department of Chemistry, Moscow State University, Moscow, Russia*

---

**Summary.** A new method of estimating the binding affinity of small organic compounds bound to biological targets is developed. The method incorporates the explicit evaluation of entropy loss for both internal and rotation-translation degrees of freedom. A small set of CK2 inhibitors with resolved crystal structures of their complexes with protein kinase CK2 was successfully fitted to experimentally derived binding constants  $K_i$ . The method is computationally efficient and accurate enough to be used in the field of high throughput receptor-oriented virtual screening.

**Keywords:** distribution functions, internal freedom degrees, virtual screening, affinity prediction.

---

**Introduction.** Efficient free energy estimation is the main challenge of virtually all molecular modeling techniques. A lot of attempts have been undertaken to cope with this task in a proper manner. Evidently, the simplest way to measure the free energy of a molecular system is the explicit numerical integration of the system's states of interest with respect to all possible configurations of the system. Monte Carlo [1, 2] and molecular dynamic simulations [3-7] are the most well-known techniques to do so. However, such direct methods are unable to deal with the configuration space of molecular systems larger than relatively small molecules with only a few degrees of freedom. Moreover, complexity grows exponentially as the number of degrees of freedom grows, which prevents any efficient handling of larger systems by direct integration even though computers and algorithms make impressive progress. As a result, more effective

algorithms of thermodynamic integration [8-14], biased potential [17-20], umbrella sampling [14-16] etc were proposed. Despite of their numerical nature such algorithms are in principle able to process huge molecular systems in reasonable time as they artificially lead their evolution through trajectories of interests. Such methods often show an excellent fit to experimental data and have become very popular in the past decades. Recent attempts have been made in the analytical estimation of free energy without explicit numerical integration. Such algorithms are definitely desirable because of natural restrictions of the numerical integration methods, but the developers of such methods run into a hardly solvable problem of the entropy loss estimation.

The most challenging application of free energy prediction is a high throughput virtual screening where additional strong demands are applied to computational complexity of algorithms used. This makes unacceptable previously mentioned high-quality methods based on numerical integration. There have been several attempts of analytical description of the free energy function that can be used in virtual screening.

---

\* Corresponding author.  
Tel./fax: +38044-5222458  
E-mail address: yakovenko.alexander@gmail.com

ening of thousands or even millions of compounds against biological targets. These attempts fail to incorporate appropriate simplifications to cope with the problem of thousands of internal degrees of freedom [21]. Most of them used artificial scoring functions mainly based on potential energy terms only [22]. Sometimes entropy is taken into account using artificial hydrophobic potentials. Finally, some of them were built on general statistical physics principles to evaluate configuration integrals. The first two groups of methods are common in molecular docking software [23-33] where the problem of internal coordinates is multiplied by the number of loops required to build all promising ligand-receptor configurations. The third group estimates entropy in three different ways. The first two are used if one deals with a small amount of well-known molecules or their parts. An example of the first approach is described in [34] where angle distributions of aminoacids were calculated as analytical functions with parameterized constant from a large set of known protein structures taken from the protein data bank [35]. Consequently, these functions were effectively used in some protein-protein docking algorithms [36], which, however, is the only field of their applicability. The second way is to formulate analytically derived distribution equations, which can be obtained for some molecular systems. Similarly to the previous ones this approach is very fast and accurate but suffers from its applicability to only one or a few molecular systems. The third approach is based on the universal treatment of generic molecular systems. If there were no approximations it would be a brilliant solution (as it is fast, accurate, and universal) for high throughput virtual screening. However significant approximations are the price to be paid for such universality.

The method presented belongs to the last subgroup of methods. As there is no known direct approach for analytical free energy estimation, all methods based on statistical physics include certain simplifications of physical laws by the introduction of certain approximations. In general, the accuracy of such methods depends rather on validity of the approximations made than on the accuracy of mathematical calculations involved. This assumption per-

mits in principle to create a method utilizing smart enough approximations to obtain good accuracy along with a significant reduction of computational complexity, which makes it applicable for use in high throughput virtual screening.

**Theory.** Following our previous work [37] we set out the formalism of configuration space measurement related to the theory of small oscillations [38], which can satisfactorily handle the problem of molecular complexes scoring. This formalism incorporates the Hessian matrix diagonalization as a good approximation to the configuration space principal axes. If otherwise stated we will use the subscript index  $e$  for rotation-translation degrees of freedom (*external*), while the subscript  $i$  stands for internal degrees of freedom. Similarly the subscripts  $b$  and  $f$  refer to the bound and the free states of the ligand, respectively. Our previously derived equation (the so called quasy-arbitrary mode) defines the free energy in terms of rotation and translation degrees of freedom:

$$\frac{Z_{eb}}{Z_{ef}} = \exp\left(3 \ln RT - 0.25 \ln D_e - \ln V + \frac{U_b}{RT}\right) \quad (1),$$

where  $Z_{eb}$  and  $Z_{ef}$  are the configuration space volumes,  $R$  is the gas constant,  $T$  is the temperature,  $D_e$  is the product of the Hessian matrix principal components (equal to its determinant),  $V$  is the volume per molecule, and  $U$  is the potential energy. Usually  $D$  is expressed in the form of the Hessian matrix determinant, which is calculated easier than matrix diagonalization. However, Eq. 1 does not account for molecular internal degrees of freedom and thus can be applied to a limited number of compounds, those which do not contain any internal degrees of freedom.

The total entropy loss can be approximated as the sum of losses in rotation-translation and in internal terms:

$$\Delta S \approx \Delta S_e + \Delta S_i \quad (2).$$

The formalism for calculating entropy of internal degrees of freedom is derived from the assumption similar to that used for deriving Eq. 1. Again, we start with the general equation for the configurations integral:

$$Z = \frac{1}{h^n} \int e^{-\beta(T(p)+U(q))} dpdq \quad (3),$$

where  $T$  is the kinetic energy,  $p$  is the momentum, and  $q$  is the coordinates.

In our calculations we use the anchor force field YFF. We deliberately designed this force field for the high accuracy treatment of the non-bonded interactions in complexes of biopolymers and small organic compounds. According to the YFF approximation torsion angles between the anchors are the only internal coordinates. As such anchors, we define a set of the largest atomic sub-graphs of a molecular graph, which does not contain any rotatable bond assigned to the atomic sub-graph edges. As the largest, we define such graph decompositions as any two anchors connected to each other correspond to a rotatable bond only.

Following the YFF formalism the internal coordinates can be represented in terms of  $m$  internal coordinates, where  $m/2$  is the number of rotatable bonds. The potential energy term is evaluated based on the theory of small oscillations:

$$Z_{bi} = e^{-\beta U_b} \left( \frac{(2\pi)^m}{\beta^m D_i} \right) \frac{1}{h^m} \int e^{-\beta T(p)} dp \quad (4).$$

In Eq. 4  $D_i$  is the product of  $m$  principal components:  $D_i = \prod_{j=1}^m b_{jj}$  of the Hessian matrix of the above defined internal. Constructing the Hessian matrix of internal coordinates of a ligand molecule is explained in detail in Appendix.

The problem with the kinetic term distribution is that we cannot reduce it similarly to the case of rotation-translation coordinates, because the mass matrix is no longer diagonal and constant. The mass matrix in the case of internal coordinates depends on the internal coordinates themselves. Using the previously calculated distribution of internal coordinates we can approximately evaluate the distribution of internal momentum:

$$\frac{Z_{bi}}{Z_{fi}} = e^{-\beta U_b} \prod_{j=1}^m \left( \left( \sqrt{\frac{1}{4\pi\beta b_{jj}}} \right) \times \left( 1 - \sqrt{\frac{1}{4\pi\beta b_{jj}}} \right) \right) \quad (5).$$

This equation arises from the assumption of the reduction of only a part of the configuration space, which corresponds to the internal coordinates accessible due to small oscillations. Consequently this part of the configuration volume should be removed from the whole kinetic energy distribution integral. However, such internal coordinate momentum treatment is reasonable for high throughput virtual screening purposes

only if the computation demand is strong and should be modified for a more serious treatment of distributions. The current approach allows reducing the problem of internal coordinate impulses distribution to the previously calculated Hessian matrix of internal coordinates.

Considering the potential energy term already accounted in Eq. 1 we obtain the analytical distribution integration equation for an arbitrary molecule:

$$K \approx e^{\beta U_b} \frac{Z_{eb} Z_{bi}}{Z_{ef} Z_{fi}} \quad (6),$$

where  $K$  is the dissociation constant of a molecular complex. It is worthwhile to mention that due to the approximations adopted, we perform the calculation of momentum distribution in Eq. 5 without any extra cost.

The evaluation of ligand — receptor interactions only is not sufficient to estimate free energy of binding: the interaction between the complex formed and the surrounding solvent molecules should be taken into account too. The desolvation energy term maybe estimated as described elsewhere. In [39] there is a new realization of the GB/SA method based on the previously developed in our labs Kirchhoff atomic charges. Here, we set out a slightly modified formalism. The break-down of the desolvation energy contribution into the polar and apolar parts continues to be the same as in the GB/SA model:

$$U_{solv} = -166 \left( 1 - \frac{1}{\epsilon} \right) \sum_{i=1}^n \left( \frac{q_i}{\alpha_i} + \sum_{\substack{j=1 \\ j \neq i}}^n \frac{q_i q_j}{f_{GB}} \right) + \sum_{i=1}^n \gamma_i A_i$$

$$f_{GB} = \sqrt{r_{ij}^2 + a_{ij}^2} \exp(-r_{ij}^2 / (4\alpha_{ij}^2)) \quad (7).$$

In Eq. 7  $\alpha_i$  is the Born radius of atom  $i$  and  $\alpha_{ij} = \sqrt{a_i a_j}$ ;  $r_{ij}$  is the distance between atoms  $i$  and  $j$ ;  $\gamma_i$  is the regression coefficient of  $i$ -th atoms type desolvation energy taken over its solvent accessible surface area (SASA);  $A_i$  is the  $i$ -th atom SASA;  $q_i$  and  $q_j$  are the atomic charges on atoms  $i$  and  $j$ , respectively;  $\epsilon$  is the solvent dielectric constant. All summations are taken over all  $n$  atoms constituting the molecule.

The modification hereby employed is the way to measure surface accessible areas and the Born radii. We implement the analytical method for rather accurate calculations of surfaces and volumes. At the first step, the Delaunay triangulation [40-42], which is a dual form of Voronoi

diagrams [43], is built. There is a problem of calculation of the surface and the volume of ball unions arose because of an arbitrary number of possible ball intersections, which makes it impossible to use the inclusion-exclusion formula [44, 45]. But many problems of calculation of multi-ball intersections can in principle be solved [45, 46]. Evidently, the Voronoi diagrams and, consequently Delaunay triangulation allow expanding the infinity of possible ball intersections into a series of up to four balls intersections. So, using the Delaunay triangulation we can obtain the general inclusion-exclusion formulas for evaluating all geometrical properties of objects in the 3D space:

$$\begin{aligned} A &= A_1 - A_2 + A_3 - A_4 \\ V &= V_1 - V_2 + V_3 - V_4 \end{aligned} \quad (8).$$

In Eq. 8 the lower index is the order of the simplex formed by regular 3D triangulations. It is well known that an arbitrary d-dimensional object can be triangulated with d+1 simplices.

Attempts of straightforward transfer of triangulation algorithms from 2D space (where this problem is perfectly solved [42]) to the 3D space were not successful. It is known that an arbitrary triangulation cannot be transformed into the Delaunay triangulation in the 3D space because of local maxima, which prevent the algorithm convergence [47]. However, an arbitrary triangulation obtained from the Delaunay triangulation, adding one new point onto an arbitrary location, can be transformed into the Delaunay triangulation in several local transformations [47]. The algorithm constructing triangulations in such an iterative way is referred to as the incremental topological flipping algorithm. The latter was proposed by Edelsbrunner [40], and here we implement it with minor changes.

Another useful result of the Edelsbrunner group is the alpha shape theory, which is a way to reduce the amount of simplices processed by the inclusion-exclusion formula (Eq. 8). One should note that even the calculation of a three balls intersection is quite a time-consuming problem, while a four balls intersection is a really hard problem. However, much calculation work can be avoided as the same absolute values often appear in the inclusion-exclusion formula, with only the signs being opposite. The alpha shapes theory allows removing such redundant work

by resorting to alpha: the balls radius growing parameter. For further detail the following references will be useful [46-51].

The generalized Born formalism has grown from the equation of single ion energy in a continuum dielectric medium. The model represented by Eq. 7 was proposed in [52] and in related works. It showed such a reasonable physical rationale that nearly all modern desolvation energy treatments are based on that idea. In the current implementation we follow this idea too. In case of multi-atomic molecule the solution is just an inverse solving for a single ion with respect to the solvent volume excluded by the other atoms in the molecule [53]. Such treatment is quite efficient due to the pair-wise additivity of the proposed evaluation scheme. Also there have been two attempts to derive efficient schemes of pair-additive Born radius calculations. The first is based on surface averaging [52, 54], and the second — on volume averaging [55] of the cells occupied by atoms. As those models were defined without any aid from the alpha shapes formalism, they are prone to errors in the prediction of geometrical properties of Voronoi cells. To handle such situations regressions coefficient was added to the models in order to provide statistical fits of the common atomic union properties. Using the alpha shapes we managed to obtain accurate excluded atomic volumes. Moreover, we can do it at a modest computation cost, because we already have the alpha shapes computed for the purpose of SASA calculations. So our evaluation of the Born radii is generally in accord with [55], but we have in hand accurate volumes and hence do not use any correction coefficients and *ad hoc* formulas:

$$\alpha_i = -166 \times \left( -\frac{166}{r_i^{vdw} + R_{solv}} + \sum_{\substack{j=1 \\ j \neq i}}^n \frac{V_j}{r_{ij}^4} \right)^{-1} \quad (9).$$

In Eq. 9  $V_j$  is the accurate volume of  $j$ -th atomic ball in the atoms union of a molecule computed from the alpha shapes theory, and  $R_{solv}$  is just the radius of the smallest ball that can be inscribed into a solvent molecule.

Summarizing all above suggestions and using Eq. 6 and 7, we define here a formula for predicting dissociation constants as follows:

$$K \approx e^{\beta(U_b - U_{solv})} \frac{Z_{eb} Z_{bi}}{Z_{ef} Z_{fi}} \quad (10).$$

**Results and discussion.** To test the method outlined we make use of the previously studied small set of resolved CK2 inhibitors with known experimental inhibition constants  $K_i$ . The first arisen issue is the adequacy of the obtained models. As we have already shown, rotation-translation distributions are not sufficient to describe experimentally derived binding constant data. The introduction of internal coordinate distribution functions as well as the prediction of desolvation energy is suggested here to treat the previously observed accuracy problems and to provide some data on ligands activity.

To establish relationships between mechanical properties of ligands and biological activity, the decomposition of ligands into rotation units is shown in Figure 1. It is interesting to trace how

the model describes internal coordinate-dependent properties of intermolecular binding. To evaluate the influence that each degree of freedom exerts on ligand binding, we use the internal coordinates Hessian matrix diagonalization technique. The Cartesian coordinates optimization was applied to obtain extreme geometry and to remove the first derivative dependencies. However, such linear energy minimization is not able to handle the whole system, and the matrix is not always positively defined. There are two sources of such problems: (i) global minima of the systems such that the coordinate vectors obtained in the course of optimization consist of small local forces in different fragments; (ii) the linear minimization algorithm is more beneficial in handling such systems than internal coordi-

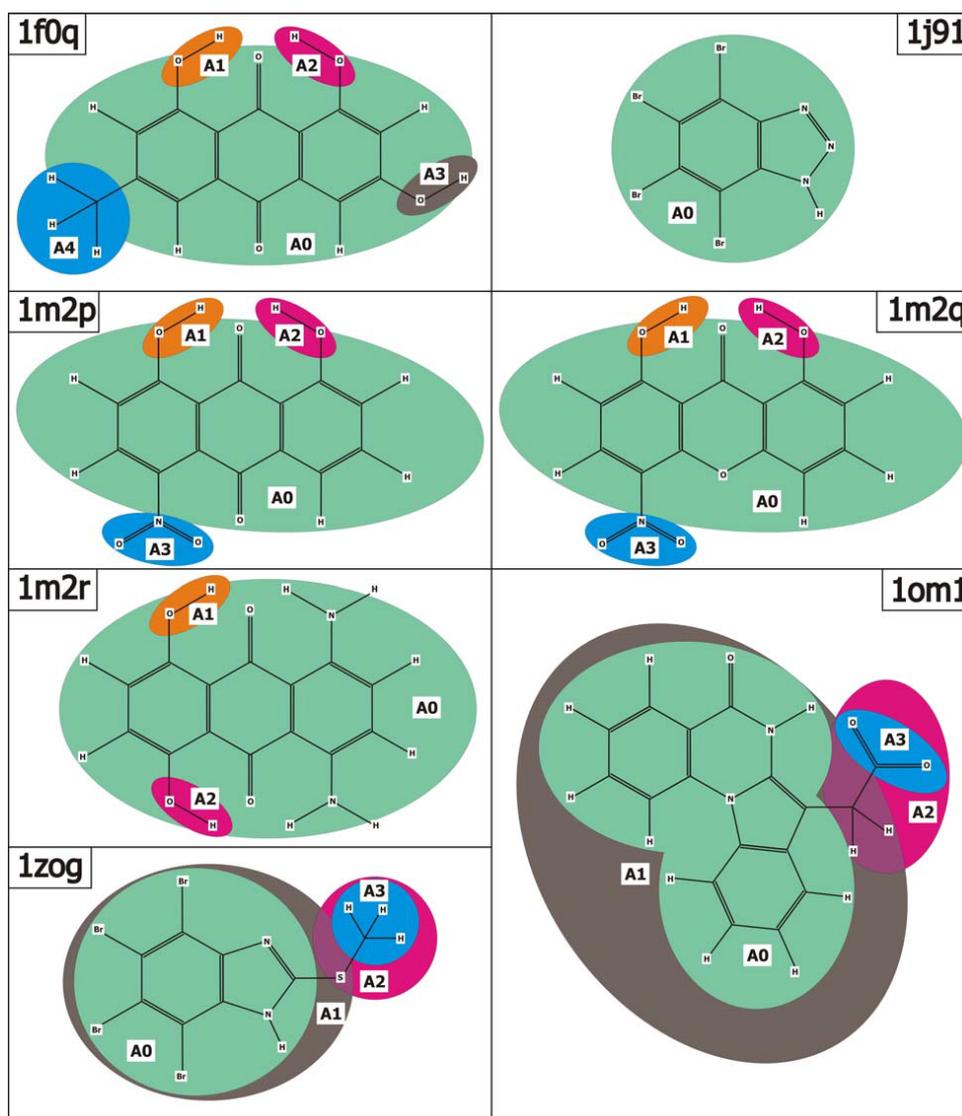


Figure 1. Decomposition of ligand internal degrees of freedom.

nate minimization engines known from the literature. As a result, diagonal elements of the Hessian matrix sometimes become negative, which corresponds to an unfavorable fragment position in the binding complex under study. To account for this effect we reduce the resulting distribution if the diagonal element is negative and multiple it if the corresponding diagonal value is positive. This simple procedure assumes that a negative fragment corresponds to the expansion of the configuration space on binding because of biasing forces arising at certain fragments in a current binding mode. The positive ones assume an ordinary collapse of the configuration space due to the complex formation. In order to evaluate the sensitivity of the models towards binding mode features we consider all complexes and their evaluations in terms of internal coordinates distributions.

A4-A0	952.6
A1-A0	-4839.5
A3-A0	5540.84
A2-A0	11116.9

**1f0q.** This complex has four rotatable bonds and only one of them feels a force bias, which results in

the negative diagonal value of the Hessian matrix and therefore the expansion of the configuration space on binding. The complex environment around the A1 anchor has a strong hydrophobic potential and the closely located methyl group of Val95. However, the water molecule-oriented A2 anchors have an opportunity to intermediate the reaction field of water and, thus, their binding is stronger than  $kT$  and is resistant to the destructive influence of A1. The A3 anchor located inside the receptor cavity provides even better collapsing of the configuration space volume because it is located in a relatively larger cavity than A1. Moreover, this cavity is more hydrophilic than A1. Taken together these factors result in a beneficial influence of the rotatable bond on the ligand potential. The methyl group A4 has a slight destructive effect as its influence on binding is lower than  $kT$ . However, the symmetry of the methyl group was not taken into account as well as relative vibrations which are three times lower comparing to the hydroxyl group. Again, the strong intermolecular hydrogen bonds at the A0, A1, and A3 anchors improve the binding effect of rotatable bonds significantly due to the significant torsion energy barrier.

**1j91.** This compound does not have any internal coordinates formed by angles between rotatable bonds and therefore, according to our formalism, its configuration space-collapsing factor of internal coordinates is set to zero.

A2-A0	-673.943
A1-A0	5638.96
A3-A0	3284.11

**1m2p.** Comparing to the previously discussed **1f0q**, the A1 and A2 anchors exert a completely opposite influence on this ligand. As in case of **1f0q** A1 anchors has significant force bias that extends its configuration volume during binding and good A2 anchors configuration space collapsing that coincide with the opposite influence that is calculated for A1. This can be explained by a different orientation of this compound regarding to the previously discussed one. The cavity of A2 is more hydrophobic, and the whole ligand is shifted to the internal hydrogen bond between residues Lys68 and Asp175. In spite of the intermolecular hydrogen bond formation, A0 and A2 still interact with those two electrostatically. Consequently, the rotatable bond with the more exposed A1 is more beneficial for the binding mode. The nitro group in A3 is significantly exposed to the strong electrostatic field from aminoacid residues Asn117, Asn118, and the backbone of Val45 as well as the Van-der-Waal's forces of the oxygen atom of A0, which corresponds to relatively smaller stability comparing to **1m2q**. The A0-A3 rotatable bond contribution, however, is estimated to be positive for binding.

A2-A0	18349.9
A1-A0	-13489.3
A3-A0	7699.4

**1m2q.** This compound is very similar to the previously discussed **1m2p** by structure and thus shows a quite similar binding mode picture. The only difference is orientation of nitro group that corresponds to the dipper burring of molecule in the kinase active site. As a result, it is better entropy losing in terms of the rotation translation entropy, inversion of A1 and A2 influences and better nitro group integration into molecular surrounding electrostatic force field. As it is concluded from calculations, the limiting factor of molecular sinking into protein medium is A1-A0 rotatable bond. This intermolecular rotor has the biggest negative value of the ones, observed in this study. At the certain level this influence compensated with A2-A0 that can stabilize intermolecular hydrogen bond in the

less dense protein core under Lys68-Asp175 ionic bond that pulls the site together. Its place from **1p2m** is now occupied with A1 that has the strongest interaction with protein and limiting future burring. The nitro group is now contributing over two times into ligands binding.

A1-A0	3882.01
A2-A0	3595.5

**1m2r.** In this complex and the two formal rotatable bonds around amines groups were reduced and the semiempirical theory level computations show a strong involvement of amines into the ring resonance system which follows with planar geometry and rotation reduction. Moreover, this group is symmetric and forms an intermolecular hydrogen bond that assumes to suppress rotation much. Accordingly with our formalism the single bonds of hydrogen bonded sp<sup>3</sup> nitrogen atoms located on the resonance system (alike amides) are not treated as rotatable. Both hydroxyl groups are located in the spacious and hydrophilic cavities as this ligand is a bit differently oriented comparing to **1m2q** and **1m2p**. Due to ligands symmetry, it does not matter interesting what exactly hydroxyl group is better located and both of them correspond nearly equally and positively onto binding.

A0-A2	2465.2
A3-A1	1665.1

**1om1.** Strong electrostatic forces around carboxyl group are close to the  $kT$  results in the positive influence of A0-A2 rotatable bond onto the binding constant. However, A3-A1 barrier is less than  $kT$  factor and thus double rotatable bonds influence ligand binding in the internal coordinates distributions negatively. The energetic contribution of carboxyl ion is huge. One should remember that this influence is not accessible without ligands groups tuning into receptor structure and all parameters should be used simultaneously in the binding description.

A2-A0	8730.1
A3-A1	3.244

**1zog.** This complex is the second example of double rotatable bonds influence onto ligands binding. Accordingly with calculations methyl fragment is quite accurately bound into receptors surrounding, however, internal rotation of methyl group is not restricted. Again certain problems with symmetry are observed and reduction in A1-A3 anchors rotations is not very significant. As for methyl group bonded through sulfur it shows a very attractive entropy loss during lig-

and binding that covers all negative effects of double rotatable bonds.

As we see from the brief analysis of the model, internal coordinates allow identifying the favorable and unfavorable rotatable bonds and establish limits in the ligands burring into receptor medium due to intermolecular rotors presents. The mixed derivative of rotation-translation freedom decreases and internal coordinates are required to be calculated for more accurate evaluations of configuration space volumes. However, here we assume negligible small connections of rotation-translation and an internal coordinates derivative similarly as it was done in [56]. The model with such approximations reproduces at last qualitative behavior of the molecular system. To evaluate the quantitative accuracy of the model we calculated the  $K_i$  according to equation 10. Before the detailed analysis of the accuracy table we have to mention about desolvation energy evaluation.

1f0q	-5.06752
1j91	-6.9555
1m2p	-5.45636
1m2q	-5.32292
1m2r	-6.4514
1om1	-7.22063
1zog	-8.18899

The desolvation energies were calculated according to the technique described in theory section. They reproduce the work amount required to transfer the ligands from low dielectric medium as protein modeled into high dielectric medium as solvent (water) modeled. This part models the relationship of protein-ligand complex and surrounding water environment. The data stored in Table are normalized by thermodynamic chaos factor  $kT$ . The most hydrophobic ligands are large compounds with a lot of hydrophobic cycles and/or halogen atoms. In the row of hydrophobic energies **1m2r** has the most negative energy that together with its binding mode, where there are no negative internal distributions observed, can explain the abnormally high activity comparing to very similar **1m2p** and **1m2q**. Note that **1f0q** has the lowest hydrophobic pressure and thus the most exposed compound and we have seen its very good internal coordinates distribution. From comparison of **1zog** with **1j91** we found that methyl-sulfur modification increases the desolvation energy significantly much better than rotation bonds reduces. Tetra-halogen fragment is nearly equivalent to additional three rings of **1om1** in desolvation energy terms.

1f0q	339.22	37.22
1j91	328.71	33.76
1m2p	335.00	34.83
1m2q	320.28	36.42
1m2r	327.17	36.57
1om1	393.40	37.40
1zog	379.05	36.10

The analysis of rotation-translation distribution shows the relations between compounds size in cubic angstroms (first column of

the table) and logarithm of the rotation-translation entropy loss (second column). The common rule is that larger compounds have bigger entropy loss during binding into the same receptors cavity. However, the reader should be warned with idea of large compound domination as they can be too big to enter site, and too flex to burry in it well, or too rigid to relax perfectly in active site. The burried area influence overcomes the common size rule for **1m2q** and **1m2r** where rotation-translation distribution is abnormally huge. As we have already seen in internal coordinates distribution section the reasons of such distribution that were unclear in our previous work now become explained in internal coordinates distributions limits. The **1f0q** compound has deeply burried methyl group that corresponds significantly to binding entropy loss.

1f0q	-5.06752	-5.73283
1j91	-6.9555	-6.39794
1m2p	-5.45636	-6.10791
1m2q	-5.32292	-6.09691
1m2r	-6.4514	-6.37675
1om1	-7.22063	-6.76955
1zog	-8.18899	-7.1549

The evaluation of method accuracy is shown in table. The predicted ligands  $K_i$  in decimal logarithmic

units are in the first column of Table 1 and experimentally derived  $K_i$  — in the second column of the article. The theoretically predicted affinities are varying more significantly comparing to experimental values. Probably better orders fit can be obtained with more accurate force field energies after its parameterization against dimers energy. In the calculation of predicted values a quasy harmonic model of rotation-transla-

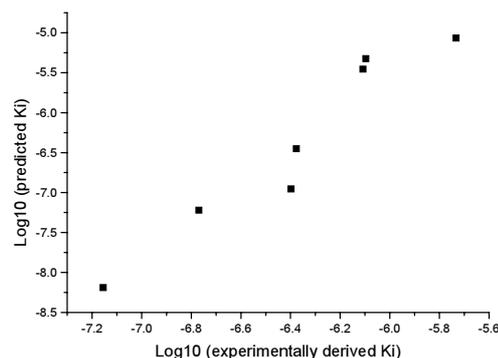


Figure 2.  $K_i$  prediction plot.

tion entropy loss was used as it was found in the previous research to be more reliable.

At last the quantitative results calculated with Eq. 10 were compared to experimentally derived  $K_i$  for the given small set. The corresponding plot is shown on Figure 2. The experimentally derived relative affinities are superposed (in decimal logarithmic scale) with theoretically predicted values. In Table 1 the errors, less than an order of magnitude, are showed in bold (green cells) and all other in single font (orange cells). The observed error does not excide the value of 1.81 magnitude order that is quite big, but the most of values are related in the range of magnitude order error.

Overall fit is excellent as no compounds are wrongly scored. Of course, such results are just a consequence of set size and in bigger sets ranging errors appear without any doubts, however, current results look quite promising for future investigations. The relation of predicted and experimental  $K_i$  showed in picture 2, and linear regression coefficient ( $r^2=0.96355$ ) calculated for this set in decimal logarithmic coordinates is quite high even despite a small set size.

**Conclusions.** The efficient method for accu-

Table 1

The quantitative methods accuracy in  $K_i$  prediction

	1f0q	1j91	1m2p	1m2q	1m2r	1om1	1zog
1f0q	0	-1.22287	<b>-0.01376</b>	<b>0.108683</b>	<b>-0.73996</b>	-1.11639	-1.69939
1j91	1.222868	0	1.209107	1.331551	<b>0.48291</b>	<b>0.106477</b>	<b>-0.47653</b>
1m2p	<b>0.013761</b>	-1.20911	0	<b>0.122444</b>	<b>-0.7262</b>	-1.10263	-1.68563
1m2q	<b>-0.10868</b>	-1.33155	<b>-0.12244</b>	0	<b>-0.84864</b>	-1.22507	-1.80808
1m2r	<b>0.739958</b>	<b>-0.48291</b>	<b>0.726197</b>	<b>0.848641</b>	0	<b>-0.37643</b>	<b>-0.95944</b>
1om1	1.116391	<b>-0.10648</b>	1.10263	1.225073	<b>0.376433</b>	0	<b>-0.583</b>
1zog	1.699393	<b>0.476526</b>	1.685632	1.808076	<b>0.959435</b>	<b>0.583003</b>	0

rate  $K_i$  estimation appropriate to the field of high throughput receptor-oriented virtual screening was developed. The method was tested on a small set of CK2 inhibitors with resolved crystal structures and experimentally derived  $K_i$ . The method evaluates entropy loss during molecular complex formation in systems of small organic compounds and huge biological polymer molecule in both translation-rotation and inter-

nal freedom decreases. In spite of a small amount of consuming computational resources the method shows good accuracy in theoretical  $K_i$  prediction from spatial 3D coordinates.

**Abbreviations.** SASA — solvent accessible surface area, SAVA — solvent inaccessible volume area.

Надійшла в редакцію 17.04.2007 р.

#### Новий метод обчислення інтегралів розподілення для задач високопродуктивного віртуального скринінгу

О.Я. Яковенко, О.О. Оліференко<sup>1</sup>, А.Г. Голуб, В.Г. Бджола, С.М. Ярмолюк

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

<sup>1</sup>Московський національний університет, Росія

**Резюме.** Розроблено метод передбачення афінності зв'язування низькомолекулярних органічних синтетичних речовин і високомолекулярних біологічних полімерів, який ґрунтується на явному обчисленні втрати ентропії як обертально-поступальних, так і внутрішніх ступенів свободи низькомолекулярних сполук. За допомогою малої тестової вибірки показано успішне відтворення обчислювальним методом експериментально визначених констант зв'язування. Розроблений метод є достатньо обчислювально ефективним і точним для застосування у високопродуктивному рецептор-орієнтованому віртуальному скринінгу баз даних синтетичних хімічних речовин.

**Ключові слова:** функції розподілення, внутрішні ступені свободи, віртуальний скринінг, передбачення афінності.

#### References

1. *Metropolis N., Ulam S.* The Monte Carlo method // *Journal of the American Statistical Association.* — 1949. — Vol. 44, No. 247. — P. 335-341.
2. *Kalos M., Whitlock P.* Monte Carlo Methods. — Wiley-Interscience, New York, 1986.
3. *Alder B.J., Wainwright T.E.* Phase transition for a hard sphere system // *J. Chem. Phys.* — 1957. — Vol. 27. — P. 1208.
4. *Alder B.J., Wainwright T.E.* Studies in molecular dynamics. I. General method // *J. Chem. Phys.* — 1959. — Vol. 31. — P. 459.
5. *Rahman A.* Correlations in the motion of atoms in liquid argon // *Phys. Rev.* — 1964. — A136. — P. 405.
6. *Stillinger F.H., Rahman A.* Propagation of sound in water. A molecular-dynamics study // *J. Chem. Phys.* — 1974. — Vol. 60. — P. 1545.
7. *McCammon J.A., Gelin B.R., Karplus M.* Dynamics of folded protein // *Nature.* — 1977. — Vol. 267. — P. 585.
8. *Beutler T.C., Mark A.E., van Schaik R.C., Gerber P.R., van Gunsteren W.F.* Avoiding singularities and numerical instabilities in free energy calculations based on molecular simulations // *Chem. Phys. Lett.* — 1994. — Vol. 222, No. 6. — P. 529-539.
9. *Pitera J.W., van Gunsteren W.F.* One-step perturbation methods for solvation free energies of polar solutes // *J. Phys. Chem. B.* — 2001. — Vol. 105, No. 45. — P. 11264-11274.
10. *Liu H., Mark A.E., van Gunsteren W.F.* Estimating the relative free energy of different molecular states with respect to a single reference state // *J. Phys. Chem.* — 1996. — Vol. 100, No. 22. — P. 9485-9494.
11. *Schafer H., van Gunsteren W.F., Mark A.E.* Estimating relative free energies from a single ensemble: Hydration free energies // *J. Comp. Chem.* — 1999. — Vol. 20, No. 15. — P. 1604-1617.
12. *Straatsma T.P., Zacharias M., McCammon J.A.* Holonomic constraint contribution to free-energy differences from thermodynamic integration molecular dynamic simulations // *Chem. Phys. Lett.* — 1992. — Vol. 196. — P. 297-302.
13. *Otter W.K.* Thermodynamic integration of the free energy along a reaction coordinate in Cartesian coordinates // *J. Chem. Phys.* — 2000. — Vol. 112. — P. 7283-7292.
14. *Sprink M., Ciccotti G.* Free energy from constrained molecular dynamic // *J. Chem. Phys.* — 1998. — Vol. 109. — P. 7737-7744.
15. *Otter W.K., Briels W.J.* The calculation of free energy from molecular dynamics with multiple constraints // *Mol. Phys.* — 2000. — Vol. 98. — P. 773-781.
16. *Otter W.K., Briels W.J.* The calculation of free differences by constrained molecular dynamics simulations // *J. Chem. Phys.* — 1998. — Vol. 109. — P. 4139-4146.
17. *Mulders A., Kruger P., Swegat W., Schlitter J.* Free energy as a potential of mean constrained force // *J. Chem. Phys.* — Vol. 104. — P. 4869-4870.
18. *Dare E., Wilson M.A., Pohorille A.* Calculating free energies using scaled-force molecular dynamics algorithm // *Mol. Sim.* — 2001.
19. *Berne B.J., Straub J.E.* Novel methods of sampling phase space in the simulation of biological systems // *Curr. Opin. Struct. Biol.* — 1997. — Vol. 7. — P. 181-189.
20. *Darve E., Pohorille A.* Calculation free energies using average force // *Annual Research Briefs.* — 2001. — P. 271-282.
21. *Teodoro M.L., Phillips G.N., Kavasaki L.E.* Molecular docking: a problem with thousands of degrees of

freedom // <http://phillipslab.biochem.wisc.edu/pdfs/108-docking.pdf>.

22. Krovat E.M., Steindl T., Langer T. Recent advances in docking and scoring // *Current Computer-Aided Drug Design*. — 2005. — Vol. 1. — P. 93-102.

23. Kuntz I.D., Blaney J.M., Oatley S.J., Langrince R., Ferrin T.E. // *J. Mol. Biol.* — 1982. — Vol. 161. — P. 269-288.

24. Rarey M., Kramer B., Lengauer T., Klebe G. A fast flexible docking method using an incremental construction algorithm // *J. Mol. Biol.* — 1996. — Vol. 261. — P. 470-489.

25. Rarey M., Kramer B., Lengauer T. Multiple automatic base selection: protein-ligand docking based on incremental construction without manual intervention // *J. Comput. Aided Mol. Des.* — 1997. — Vol. 11. — P. 369-384.

26. Goodsell D.S., Olson A.J. Automated docking of substrates to proteins by simulated annealing // *Proteins*. — 1990. — Vol. 8. — P. 195-202.

27. Trosset J.Y., Scheraga H.A. PRODOCK: software package for protein modeling and docking // *J. Comput. Chem.* — 1999. — Vol. 20. — P. 412-427.

28. Trosset J.Y., Scheraga H.A. Reaching the global minimum in docking simulations: a Monte Carlo energy minimization approach using Bezier splines // *Proc. Natl. Acad. Sci. USA*. — 1998. — Vol. 95. — P. 8011-8015.

29. Trosset J.Y., Scheraga H.A. Flexible docking simulations: Scaled collective variable Monte Carlo minimization approach using Bezier splines, and comparison with a standard Monte Carlo algorithm // *J. Comput. Chem.* — 1999. — Vol. 20. — P. 244-252.

30. Jones G., Willett P., Glen R.C., Leach A.R., Taylor R. Development and validation of a genetic algorithm for flexible docking // *J. Mol. Biol.* — 1997. — Vol. 267. — P. 727-748.

31. Jones G., Willett P., Glen R.C. A genetic algorithm for flexible molecular overlay and pharmacophore elucidation // *J. Comput. Aided Mol. Des.* — 1995. — Vol. 9. — P. 532-549.

32. Boehm H.J. LUDI: rule-based automatic design of new substituents for enzyme inhibitor leads // *J. Comput. Aided Mol. Des.* — 1992. — Vol. 6. — P. 593-606.

33. Boehm H.J. The computer program LUDI: a new method for de novo design of enzyme inhibitors // *J. Comput. Aided Mol. Des.* — 1992. — Vol. 6. — P. 61-78.

34. Abagyan R., Totrov M. Biased probability Monte Carlo conformation searches and electrostatic calculations for peptides and proteins // *J. Mol. Biol.* — 1994. — Vol. 235. — P. 983-1002.

35. Berman H.M., Westbrook J., Feng Z., Gilliland G., Bhat T.N., Weissig H., Shindyalov I.N., Bourne P.E. The Protein Data Bank // *Nucleic Acids Research*. — 2000. — Vol. 28. — P. 235-242.

36. Cavasotto C.N., Orry A.J.W., Abagyan R. The challenge of considering receptor flexibility in ligand docking and virtual screening // *Curr. Comput. Aided Drug Des.* — 2005. — Vol. 1. — P. 423-440.

37. Yakovenko O.Ya., Golub A.G., Bdzholo V.G., Yarmoluk S.M. Application of distribution function of rotation and translation degrees of freedom for CK2 inhibitors K<sub>i</sub> estimation // *Ukr. Bioorg. Acta*. — 2006. — Vol. 4, No. 2. — P. 47-55.

38. Landau L.D., Lifshits E.M. *Teoreticheskaja Fizika*. — M.: Nauka, 1988.

39. Rizzo R.C., Aynechi T., Case D.A., Kuntz I.D. Estimation of absolute free energies of hydration using continuum methods: accuracy of partial charge models and optimization of nonpolar contributions // *J. Chem. Theory Comput.* — 2006. — Vol. 2. — P. 128-139.

40. Edelsbrunner H., Shah N.R. Incremental topological flipping works for regular triangulations // *Algoritmica*. — 1996. — Vol. 15. — P. 223-241.

41. Borouchaki H., Lo S. Fast Delaunay triangulation in three dimensions // *Comput. J. Numer. Methods Engrg.* — 1995. — Vol. 128. — P. 153-167.

42. Skvotsov A.V. Survey of algorithms for Delaunay triangulations constructions // *Numerical Methods and Programming*. — 2002. — Vol. 3. — P. 14-39.

43. Watson D.F. Computing the n-dimensional Delaunay tessellation with applications to Voronoi polytopes // *Comput. J.* — 1981. — Vol. 24. — P. 167-172.

44. Liang J., Edelsbrunner H., Fu P., Sudhakar P.V., Subramanian S. Analytical shape computations of molecules: I. Molecular area and volume through alpha shape // *Proteins*. — 1998. — Vol. 33. — P. 1-17.

45. Dodd L.R., Theodorou D.N. Analytical treatment of the volume and surface area of molecules formed by an arbitrary collection of unequal spheres intersected by planes // *Molecular Physics*. — 1991. — Vol. 72, No. 6. — P. 1313-1345.

46. Edelsbrunner H., Fu P. Measuring space filling diagrams and voids, report to Nacional Science Foundation under grant ASC-9200301.

47. Galanin M.P., Sheglov I.A. Development and Implementation of Algorithms for Constrained Volume Triangulations: Iterative Algorithms Preprint, Inst. Appl. Math., the Russian Academy of Science, 2006.

48. Edelsbrunner H., Mucke E.P. Simulation of simplicity: a techniqu to cope with degenerate cases in geometric algorithms // *Transactions on Graphics*. — 1990. — Vol. 9, No. 1. — P. 66-104.

49. Edelsbrunner H., Kirkpatrick D.G., Seidel R. On the shape of a set of points in the plane *IEEE transactions on information theory*. — 1983. — Vol. 29, No. 4. — P. 551-559.

50. Edelsbrunner H. The union of balls and its dual shape National Science Foundation report under grant ASC-9200301.

51. Edelsbrunner H. Weighted alpha spheres.

52. Hawkins G.D., Christopher J.C., Donald G.T. Parameterized model of aqueous free energies of solvation based on pairwise descreening of solute atomic charges from a dielectric Medium // *J. Phys. Chem.* — 1996. — Vol. 100. — P. 19824-19839.

53. Born M.Z. // *Phys.* — 1920. — Vol. 1. — P. 45.

54. Hawkins G.D., Cramer C.J., Truhlar D.G. Pairwise solute descreening of solute charges from dielectric medium // *Chem. Phys. Lett.* — 1995. — Vol. 246. — P. 122-129.

55. Qiu D., Shenkin P.S., Hollinger F.P., Still W.C. The GB/SA continuum model for solvation. A fast analytical method for the calculation of approximate Born radii // *J. Phys. Chem.* — 1997. — Vol. 101. — P. 3005-3014.

56. Schafer H., Mark A.E., van Gunsteren W.F. Estimating relative free energies from a single ensemble: Hydration free energies // *J. Chem. Phys.* — 2000. — Vol. 113. — P. 7809-7817.

## APPENDIX

**Constructing the internal coordinates  
Hessian matrix**

There are two cases of anchors interactions, observed in internal coordinates Hessian matrix formation. The first is just a relative rotation of one anchor respectively to another. In this case the second energy derivative is taken as the sum of second derivatives of atoms in different anchors interaction. The second case is the rotations of ligands fragment relatively to an other system around two rotatable bonds simultaneously. This kind of rotation can be easy treated similarly to the first case via bonding the coordinates origin with fragment located between two rotatable bonds. In this case the problem is solved using the first case formalism, and the summation is performed over all system but anchors located between two rotatable bonds. So we just need to handle the case of relative rotation of two anchors each around its own rotatable bond. Lets consider Figure 3, where this situation is described.

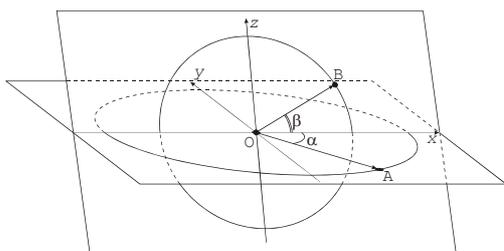


Figure 3. Rotation planes superposition.

Described situation is observed after superposing rotation centers of both anchors. The abscissa axis is the line of rotation planes intersection, the applicate axis is the rotation axis of the first anchor and the axis of ordinates is formed as a cross-product of the previous axis. The vector of the abscissa axis is calculated as cross-product of two rotation axes of anchors A and B. The angles  $\alpha$  and  $\beta$  are computed as cosines of angles formed by the abscissa axis with OA and OB vectors, respectively. The lengths of OA and OB vectors are denoted as  $r_\alpha$  and  $r_\beta$ . The square distance between points A and B can be easy calculated as the sum of squared projections lengths. Taking into account that these anchors were moved previously by the vector between their rotation centers denoted as  $r_\gamma$  and keeping in mind the angle  $\psi$  between two planes calculated as an angle between their rotation axes we obtain the equation of distance measurement bet-

ween points A and B as functions of rotation angles  $\alpha$  and  $\beta$ :

$$r_{AB}^2(\alpha, \beta) = \left( r_\alpha \cos\alpha - r_\beta \cos\beta + \frac{\vec{r}_\gamma \cdot \vec{r}_o}{|\vec{r}_o|} \right)^2 + \left( r_\alpha \sin\alpha - r_\beta \sin\beta + \frac{\vec{r}_\gamma \cdot \vec{y}_o}{|\vec{y}_o|} \right)^2 + \left( r_\alpha \sin\alpha - r_\beta \sin\beta + \frac{\vec{r}_\gamma \cdot \vec{y}_o}{|\vec{y}_o|} \right)^2 \quad (10).$$

The case of diagonal elements of Hessian matrix evaluation the much simpler formalism can be considered. This case is depicted in Figure 4.

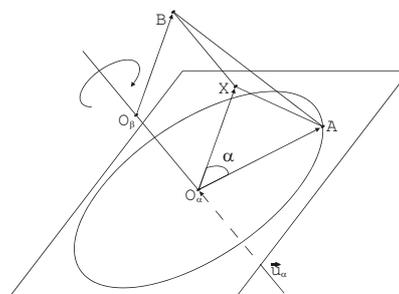


Figure 4. The diagonal element of internal coordinates Hessian matrix scheme.

Suggesting point B is the position of atom from system fragment that does not belong to the rotation group of current internal freedom degree. Point A is the atom that belongs to the fragment of the system that is registered in rotation group of current internal freedom degree and  $u_\alpha$  is the vector of rotation axis. Then, the distance AB can be defined as the function of rotation angle  $\alpha$ . The cosines of angle  $\alpha$  value is calculated via scalar product of vectors  $O_\beta B$  and  $O_\alpha A$ . The distance is expressed in the form of:

$$r^2 = AB^2 = XA^2 + XB^2 = XB^2 + O_\beta B^2 + O_\alpha A^2 - 2O_\beta B \cdot O_\alpha A \cdot \cos\alpha = C - K \cos\alpha \quad (11),$$

where  $C$  is a certain constant that eliminates after differentiating, and  $K$  is a double product of rotation radii of point A and B relatively to  $u_\alpha$  axis. The second derivatives of distance square over rotation angles in Eq. 10 and 11 are rather evident and not shown.

The most of today force fields YFF atomic interactions potential depends on the distances between interaction centers variables only. All the other parameters are just parameterized constants and having Eq. 10 and 11 one can easy evaluate the Hessian matrix of internal coordinates. Just mention that derivative over square radius as defined above should be accounted in energy derivative as:

$$H_{ij} = \frac{1}{4r^2} \left( \frac{d^2U}{dr^2} - \frac{1}{r} \frac{dU}{dr} \right) \frac{\partial r^2}{\partial_j} \frac{\partial r^2}{\partial_i} + \frac{1}{2r} \frac{dU}{dr} \frac{\partial^2 r^2}{\partial_j \partial_i}$$