ПРОГРАММИРОВАНИЕ

INFLUENCE OF THE METHOD OF HYDROGEN ATOMS INCORPORATION INTO THE TARGET PROTEIN ON THE PROTEIN-LIGAND BINDING ENERGY

D.C. Kutov, E.V. Katkova, A.V. Sulimov, O.A. Kondakova, V.B. Sulimov
Dimonta Ltd., Moscow, Russian Federation
Research Computer Center, Lomonosov Moscow State University, Moscow,
Russian Federation
E-mail: dk@dimonta.com, katkova@dimonta.com, as@dimonta.com, ok@dimonta.com,
vs@dimonta.com

Preparation of the target-protein, particularly the protein protonation method can affect considerably the spatial arrangement of the attached hydrogen atoms and the charge state of individual molecular groups in amino acid residues. This means that the calculated protein-ligand binding energies can vary significantly depending on the method of the protein preparation, and it also can lead to the different docked positions of the ligand in the case of docking (positioning of the ligand in the protein active site). This work investigates the effect of the hydrogen atoms arrangement method in the target-protein on the protein-ligand binding energy. All hydrogen atoms of target-protein are fixed or movable. The comparison of the protein-ligand binding energies obtained for the test set of target-proteins prepared using six different programs is performed and it is shown that the protein-ligand binding energy depends significantly on the method of hydrogen atoms incorporation, and differences can reach 100 kcal/mol. It is also shown that taking into account solvent in the frame of one of the two continuum implicit models smooths out these differences, but they are still about $10-20$ kcal/mol. Moreover, we carried out the docking of the crystallized (native) ligands from the protein-ligand complexes using the SOL program and showed that the different methods of the hydrogen atoms addition to the protein can give significantly different results both for the positioning of the native ligand and for its protein-ligand binding energy.

Keywords: target-protein; crystal structure; protonation; protein-ligand binding energy; docking; force field.

Introduction

Effectiveness of molecular modelling application to development of new drugs [1] depends on many factors. One of such important factors is quality of the three-dimensional model of the target protein structure. Commonly, hydrogen atoms are missed in protein structures taken from the Protein Data Bank [2,3]. So, it is necessary to add hydrogen atoms to a protein structure by any suitable program. Different methods of hydrogen atoms addition to the protein are used in programs of this kind. Not only different algorithms can be used in such programs but also the different methods of molecular modelling affecting both a spatial arrangement of added hydrogen atoms and charge states of some amino acids molecular groups. The aim of this paper is to investigate the influence of hydrogen atoms incorporation method on the protein-ligand binding energy when

the energy is calculated using the crystallized (native) ligand position from the protein-ligand complex. A comparison of protein-ligand binding energies obtained for a test set of protein-ligand complexes prepared by six different programs is fulfilled. Essential dependence of the binding energy on the method of hydrogen atoms incorporation to the target protein is demonstrated. Strong influence of mobility of target protein hydrogen atoms on the protein-ligand binding energy is demonstrated. The influence of solvent models on variations of the protein-ligand binding energy with the technique of hydrogen atoms incorporation is also investigated. Two different implicit solvent models are used for this purpose. Besides, the protein preparation technique influences strongly docking results as well. Docking of native ligands in their own proteins prepared by different hydrogen atoms incorporation methods is carried out by docking program SOL [4–6]. Protein-ligand binding energies and ligand positions after docking are analyzed for all the protein preparation techniques.

1. Methods

1.1. A Test Set of Protein-Ligand Complexes

The calculations are carried out for the test set of 16 protein-ligand complexes [7], taken from the Protein Data Bank (PDB) [2, 3]. Only the complexes with high X-Ray resolution (less than 2 Å) were selected. Some of the complexes are the same proteins co-crystallized with different ligands, the other ones are structures of different proteins co-crystallized with their native ligands. The structure of low-molecular ligands in the test set varies from small rigid molecules (4FSW, 26 atoms included hydrogen atoms) up to large molecules with significant conformational mobility (1VJ9, 74 atoms including hydrogen atoms). Here is the full list of complexes:

- 4 complexes of CHK1 (checkpoint kinase 1) protein (4FT0, 4FT9, 4FSW, 4FTA);
- 2 complexes of ERK2 (extracellular signal-regulated kinase 2) protein (4FV5, 4FV6);
- 2 complexes of thrombin protein (1DWC, 1TOM);
- 6 complexes of urokinase protein (1C5Y, 1F5L, 1O3P, 1SQO, 1VJ9, 1VJA);
- 2 complexes of factor Xa protein (2P94, 3CEN).

1.2. Hydrogen Atoms Incorporation into a Target Protein and a Ligand

Initially, all atoms and molecules with "HETATM" notation, including a ligand, were removed from the file with the PDB complex structure. Then the next programs, having the functional of hydrogen atoms addition, were applied to prepared structures:

1. The APLITE program [6, 8] adds hydrogen atoms according to the standard protonation states of amino acids at pH = 7,4. Protonation states of hisitidines were selected by the comparison of hydrogen atoms electrochemical potentials in "HD1" and "HE2" positions. Then APLITE optimizes the protein energy in the frame of MMFF94 force field in respect to all hydrogen atoms positions. All possible positions of moved hydrogen atoms are examined during the optimization, for example, hydrogen positions of tyrosine's hydroxyl group. During the optimization protein heavy atoms (excluded hydrogens) are fixed at their positions corresponding to the PDB coordinates.

2. The MacroModel program (Schrodinger Release 2015-3: Schrodinger Suite 2015-3 Protein Preparation Wizard; Epik version 3.3, Schrodinger, LLC, N.Y., 2015; Impact version 6.8, Schrodinger, LLC, N.Y., 2015; Prime version 4.1, Schrodinger, LLC, N.Y., 2015) performs optimization of the protein energy in respect with variations of hydrogen atoms positions [9]. The MacroModel is designated as MModel in the following tables. Initially hydrogen atoms are added to any heavy protein atom, where it is necessary, by the Preprocess procedure and the subroutine applyhtreat. Then the hydrogen atoms positions are optimized in the frame of OPLS 2005 force field by an automatic mode in particular by a re-orientation of hydroxyl and amide groups in Asn and Gln amino acids. The charge states of Asp and Glu amino acids as well as the His charge state and position are also determined during the optimization.

3. The program Chimera 1.10.1 [10] (different methods realized in this program are designated as Chimera1 or Chimera 2 in the following) uses atom and amino acid names to determine an atom type and consequently the number of the required hydrogen atoms for this atom type. On default Glu, Asp amino acids and $C-terminal$ carboxyl group have negative charges, Lys, Arg and $N-terminal$ amino group have positive charges. The protonation of histidines depends on its state (Hid, Hie, Hip) the method used and the local environment in case of His notation. Two techniques of hydrogen atoms addition are realized in the program. First (i), hydrogen atoms are added according to atom types to escape collisions with other atoms. This technique considers the steric limitation only and known as "steric only" (Chimera 1). Second (ii), hydrogen atoms are added according with atom types to escape collisions with other atoms and to form hydrogen bonds where possible. This method is called here as Chimera 2. The energy minimization is not performed. Both methods are used to add hydrogen atoms to protein structures in this paper.

4. The program Reduce 3.23.130521 [11] (further it is designated as Reduce) adds hydrogen atoms according to valency and atom types. Amino groups and carboxyl groups are ionized (NH_3^+ and COO^- correspondingly), histidines have no ionized state on default. The program optimizes an orientation of such chemical groups as OH^-, SH^-, NH_3^+, $methyl$, amide group of Asp, Glu side chains, histidine rings as well.

5. The program AutodockTools 1.5.6 [12] (further it is designated as AutodockTools) adds hydrogen atoms by the procedure based on the Babel program [13]. At this technique amino and carboxyl groups, Arg and His amino acids are ionized, with the corresponding charges NH_3^+, COO^-, Arg^+ and His^+. There is no optimization of attached hydrogen atoms.

6. The MOPAC2016 program [14] (MOPAC2016, Stewart J.J.P., Stewart Computational Chemistry, Colorado Springs, CO, USA, Available at: http://OpenMOPAC.net) adds hydrogen atoms according to valence of protein heavy atoms converting the protein to the neutral state. The MOPAC2016 is designated as Mopac in the following tables. Then amino and carboxyl groups as well as Arg amino acids are ionized up to the charge states NH_3^+, COO^- and Arg^+, correspondingly. The next step is the protein energy optimization with respect to positions of all hydrogen atoms by the LBFGS (Limited-memory Broyden $-$

Fletcher – Goldfar – Shanno) algorithm when the protein energy is calculated in the frame of the new semiempirical quantum chemical method PM7 [14].

An initial ligand conformation was obtained from the crystalline structure of the corresponding PDB complex. Hydrogen atoms were added to the ligand by the Avogadro program [15, 16] at pH = 7,4.

Different methods of hydrogen atoms incorporation create in some cases different protein charge states but the ligand charge state is one and the same for a given complex because it is defined by the Avogadro program. Table 1 demonstrates total protein charges obtained by different methods of protein preparation as well as total ligand charges. It is obvious from Table 1 that APLITE, MacroModel, Chimera 2 and MOPAC create the identical protonation of amino acid residues almost in all cases. This results in equal charges of the same proteins. At the same time Chimera 1 and Autodock carried out a completely different protonation. Protein charges after Chimera 1 and Autodock application are very different from one another and from the charges obtained after application of the programs from the first group. Reduce falls in grey area in this sense: this program does the same protonation as the programs from the first group do for approximately half of the complexes.

Table 1

The total protein charges for six different methods of hydrogen atoms incorporation. The total ligand charges are created by the Avogadro program

PDB id	Protein charges							Ligand charges
	APL	MM	Chi1	Chi2	Rdc	ADT	Mop	Avogadro
1C5Y	2	2	11	2	2	10	2	1
1DWC	−3	−3	2	−3	−4	8	−3	0
1F5L	3	3	12	3	3	17	3	1
1O3P	4	4	13	4	4	12	2	1
1SQO	3	3	12	3	3	18	3	1
1TOM	4	4	9	4	2	13	4	2
1VJ9	3	3	12	3	3	19	3	1
1VJA	3	3	12	3	3	20	3	1
2P94	1	1	6	1	1	4	1	0
3CEN	1	1	6	1	1	7	1	0
4FSW	−1	−1	6	−1	−3	2	−1	0
4FT0	−6	−6	1	−6	−7	−4	−6	−1
4FT9	−3	−3	4	−3	−5	0	−1	0
4FTA	−1	−1	6	−1	−3	2	−1	−1
4FV5	−2	−2	9	−2	−4	5	−2	0
4FV6	−1	−1	10	−1	−3	8	2	0

* Only for tables in this article all programs for the hydrogen atoms incorporation have short names: Aplite – APL, MModel – MM, Chimera 1 – Chi1, Chimera 2 – Chi2, Reduce – Rdc, AutoDockTools – ADT, Mopac – Mop.

2. Results and Discussion

2.1. The Calculation of Protein-Ligand Binding Energy

The protein-ligand binding energy ΔG was calculated as follows:

$$\Delta G = G(PL) - G(P) - G(L), \tag{1}$$

where $G(PL)$, $G(P)$ and $G(L)$ are energies of the protein-ligand complex, the protein and the ligand, respectively calculated in the frame of MMFF94 force field [17] in vacuum.

In the present work we consider only the enthalpy component of the binding energy without taking into account the entropy contribution. It is enough to estimate the dependence of the calculated protein-ligand binding energy on the method of hydrogen atoms incorporation into the protein structure. Positions of protein heavy atoms in equation (1) correspond to respective PDB [2, 3] data. Hydrogen atoms positions are obtained by one of the six programs described above. The free ligand conformation in equation (1) remains the same for all hydrogen atoms incorporation methods used for the protein preparation. This conformation corresponds to the energy global minimum in vacuum in MMFF94 force field found by the FLM program [7, 8]. The energy of the protein-ligand complex in (1) is equal to the energy local minimum in vacuum in MMFF94 force field when the energy optimization is performed by LBFGS method starting from the native ligand pose in the crystallized complex structure. The energy optimization is performed in respect with all ligand atoms positions while positions of all protein atoms are kept fixed.

The case of all movable protein hydrogen atoms is also considered. In this case the configuration of the protein in equation (1) corresponds to the energy local minimum of the protein in the MMFF94 force field in vacuum. The local optimization is performed in respect with positions of all protein hydrogen atoms. In this case (all hydrogen atoms of the protein are movable) the optimization of the protein-ligand complex energy is carried out with respect to positions of all ligand atoms and also positions of all protein hydrogen atoms. The root means square deviations (RMSD) between the optimized native ligand and the crystallized native ligand positions in the complex were also compared for two cases: for movable and fixed hydrogen atoms.

Thus estimating the protein-ligand binding energy we neglect the solvent contribution and also we use the local energy minimum of the complex instead of employment of the global energy minimum.

Obviously the conclusions made on the basis of such calculations about the influence of the hydrogen atoms incorporation method should be applicable to the case when the configurations of the complex and the protein in equation (1) correspond to the global energy minimum obtained with variations of coordinates of respective atoms.

For several hydrogen incorporation methods the binding energy is calculated in vacuum and in solvent both. Interaction with water solvent is described in the frame of two implicit (continual) models: the Surface Generalized Born model (SGB) [18] implemented in the program DISOLV [19] and the Polarized Continuum Model (PCM) [20,21]. The complex and the ligand conformations obtained during the vacuum optimization are used for the energy calculation in solvent without their additional optimization. The protein conformation loaded from the PDB and prepared by the respective program (hydrogen atoms incorporation) is used for the protein energy calculation in solvent.

Different methods of hydrogen atoms incorporation into the protein structure supposedly can influence the protein-ligand binding energy and the ligand position both after the docking procedure. The docking program SOL [4–6] is used to check this assumption. The SOL program determines the global minimum of the protein-ligand binding energy in the MMFF94 force field by the genetic algorithm. The protein conformation is rigid during the docking procedure but the ligand is flexible. The ligand can change its conformation by torsion rotations, as well as by rotations and translations as a whole.

2.2. Results and Discussion

The protein-ligand binding energies corresponding to the protein models prepared by different hydrogen atoms incorporation methods are presented in Table 2.

The binding energy ΔG in vacuum strongly depends on the hydrogen atoms incorporation method as it can be seen from Table 2. The binding energy differences can reach 100 kcal/mol for different methods of hydrogen atoms incorporation. Such large energy differences first of all are defined by differences of total protein charges. Total protein charge is the sum of individual amino acids charges and it depends on the hydrogen atoms incorporation method. For example, we can compare the values of the binding energy in "APLITE" and "AutoDock" columns of Table 2.

If we compare only results for proteins prepared by different methods but so that the same amino acids residues are charged equally and respectively total protein charges are the same (this situation occurs when APLITE, MacroModel and Chimera 2 programs are used, and when MOPAC is used for most of the complexes) then differences in binding energies caused by only different hydrogen atoms positions are not so large and they are equal to several kcal/mol for the half of the complexes under consideration (for example, for 2P94 complex). At the same time there are complexes with the energy difference of about dozens of kcal/mol. For example, the binding energies of 1SQO complex prepared by Chimera 2 and APLITE differ on 50 kcal/mol. This large difference in binding energies (for the same complex) obviously is due to only the difference in the spatial arrangement of hydrogen atoms added to the protein by different methods.

Search for the energy minimum of the complex and the free protein taking into account the mobility of the hydrogen atoms of the protein gives the same picture (Table 2). In this case the binding energies can change significantly depending on a given protein and a specific method of hydrogen atoms incorporation (comparing with the case when all protein atoms are fixed): from a few kcal/mol to more than 25 kcal/mol. Accounting of the mobility of the protein hydrogen atoms during energy optimization of the complex does not lead to the essential changes in their positions. For example, for the 1O3P complex (prepared using APLITE) RMSD between protein hydrogen atoms positions before and after optimization with movable hydrogen atoms is only RMSD $\approx 0{,}06$ Å, though a few hydrogen atoms have maximal shift (1,7 Å) and a few dozen hydrogen atoms have noticeable shift of about 0,5 Å.

There are significant differences in the spatial arrangement of the hydrogen atoms added by different methods for the same protein structure. RMSD values between hydrogen atoms added with APLITE and with all other programs are about 0,7 Å for all considered complexes. These roots mean square deviations consisting of large distances (1,6 – 1,7 Å) for the substantial part (over a quarter) of all protein hydrogen atoms and the tenths of angstrom for the rest of the hydrogen atoms. For example, for the 1O3P com-

The protein-ligand binding energy ΔG in kcal/mol, calculated
by equation (1) in MMFF94 force field in vacuum for the different methods of hydrogen
atoms incorporation into the protein. For each protein-ligand complex (the first column
from the right) there are two rows in the table: the upper row shows the binding energy
with fixed protein hydrogen atoms and the lower one shows the binding energy
calculated for moveable protein hydrogen atoms

PDB id	H fixed/mov	ΔG, kcal/mol						
		APL*	MM	Chi1	Chi2	Rdc	ADT	Mop
1C5Y	H_{fixed}	−113,5	−152,9	15,4	−151,7	−154,7	1,6	−146,5
	H_{mov}	−119,7	−156,2	18,1	−142,5	−133,6	−19,7	−146,1
1DWC	H_{fixed}	−57,4	−68,2	−90,9	−82,7	−91,5	−115,6	−61,9
	H_{mov}	−62,6	−64,7	−98,4	−91,1	−89,6	−95,8	−54,7
1F5L	H_{fixed}	−98,6	−119,6	45,6	−134,8	−132,3	120,5	−100,6
	H_{mov}	−113,7	−115,8	62,3	−113,6	−112,9	142,9	−107,7
1O3P	H_{fixed}	−114,3	−148,4	32,0	−154,8	−148,0	21,3	−165,0
	H_{mov}	−118,1	−136,2	42,3	−133,1	−123,4	25,0	−169,6
1SQO	H_{fixed}	−98,3	−115,0	41,4	−149,0	−149,5	115,4	−114,9
	H_{mov}	−121,1	−127,2	29,6	−131,1	−132,7	132,0	−126,1
1TOM	H_{fixed}	−147,1	−162,8	20,0	−177,6	−269,1	115,8	−144,1
	H_{mov}	−151,3	−148,3	38,3	−149,5	−241,3	130,6	−140,7
1VJ9	H_{fixed}	−133,2	−166,2	4,8	−180,2	−183,6	93,7	−138,9
	H_{mov}	−157,3	−155,2	22,8	−152,7	−145,2	112,4	−155,3
1VJA	H_{fixed}	−134,4	−166,0	18,2	−186,6	−171,1	134,6	−154,8
	H_{mov}	−161,0	−160,0	39,9	−163,6	−151,1	144,9	−163,3
2P94	H_{fixed}	−43,1	−46,7	−56,4	−49,2	−53,0	−84,1	−43,1
	H_{mov}	−44,7	−47,9	−50,2	−46,8	−49,4	−86,4	−35,2
3CEN	H_{fixed}	−49,7	−51,0	−49,9	−52,0	−54,8	−76,6	−49,0
	H_{mov}	−51,2	−50,4	−49,5	−49,8	−55,9	−74,2	−56,4
4FSW	H_{fixed}	−43,8	−42,0	−43,9	−43,2	−48,4	−50,2	−44,2
	H_{mov}	−44,4	−43,4	−42,3	−43,4	−44,6	−50,9	−45,4
4FT0	H_{fixed}	103,5	104,1	−18,0	101,1	122,4	93,6	98,7
	H_{mov}	88,1	94,3	−24,3	87,2	115,8	78,8	92,7
4FT9	H_{fixed}	−47,9	−45,6	−48,9	−47,2	−46,3	−47,5	−45,5
	H_{mov}	−49,0	−47,8	−41,1	−47,9	−46,3	−48,2	−28,3
4FTA	H_{fixed}	89,5	88,8	11,5	95,1	118,1	60,1	93,4
	H_{mov}	80,5	85,8	1,2	82,6	106,7	56,0	85,5
4FV5	H_{fixed}	−60,9	−65,1	−59,3	−63,4	−71,6	−59,6	−61,6
	H_{mov}	−62,4	−64,3	−60,8	−64,9	−66,5	−59,3	−64,6
4FV6	H_{fixed}	−71,8	−77,0	−72,1	−76,3	−82,9	−75,1	−72,8
	H_{mov}	−74,9	−77,0	−74,3	−77,2	−79,0	−73,7	−74,1

* How to read the short program names, shown in the Table 1.

plex we compare positions of the hydrogen atoms added using the APLITE and MOPAC programs and we find that distances between 503 hydrogen atoms (from 1893 added hydrogen atoms) are 1,1 – 1,8 Å, and ones between other protein hydrogen atoms are less than 0,1 Å. The calculation of RMSD between the atoms of the optimized crystallized ligand and the atoms of the initial ligand pose in the crystal complex are performed. It is found that none of the ligand is displaced more than 2 Å from the initial crystallized pose after optimization. The exception is the 4FSW complex, which was prepared using "Autodock-Tools" (RMSD > 5 Å for proteins with movable and fixed hydrogen atoms) and "Reduce" (RMSD = 2,9 Å for the protein with fixed hydrogen atoms, but RMSD = 0,7 Å when hydrogen atoms mobility is taken into account).

On the whole, accounting mobility of protein hydrogen atoms during the protein-ligand energy optimization does not result in a large change of the optimized ligand position comparing with the case of the complex energy optimization with fixed protein hydrogen atoms: respective values of RMSD (between the initial and optimized ligand poses) differ from one another on not more than 0,1 Å .

Analysis of Table 2 demonstrates that the binding energies for proteins prepared by Chimera 1 and AutoDockTools differ significantly from ones for proteins prepared by other five methods. It is not surprisingly because total protein charges obtained with these two methods are quite different from one another and from protein total charges obtained by other methods. We calculated pairwise correlation coefficients of the binding energies for all considered hydrogen atoms incorporation methods and can conclude the next:

1. The methods included in programs APLITE, MacroModel, Chimera 2, Reduce and Mopac have the perfect pairwise correlations (correlation coefficients are equal to 0,96 – 0,99).

2. The methods included in programs Chimera 1 и AutoDockTools correlate with one another rather well (correlation coefficient is equal to 0,87).

3. Any correlation between the two groups of methods is absent. These results are also connected with the difference in protonation states of distinct molecular groups and therefore with the difference in total protein charges.

It can be assumed that so large difference of binding energies for different hydrogen atoms positions is determined by the solvent (water) exclusion from the calculation. The solvent (water) screens strongly Coulomb interactions of atom charges in a molecular system owing to high water permittivity (78,5 at 300 K). To verify this assumption binding energy calculations for the several hydrogen atoms incorporation methods are accomplished with solvent using two implicit water models. Respective results are presented in Table 3.

It can be seen from Table 3 that solvent accounting strongly reduces values of protein-ligand binding energies. Obviously, first of all it is the consequence of screening of atoms Coulomb interactions by solvent polarized charges, i.e. the desolvation effect [18–20]. The binding energy difference for the two different solvent models is insignificant for the majority of the complexes and is equal to several kcal/mol (see the Table 3), although for some complexes this difference can reach 10 kcal/mol. So, it is possible to use the SGB [18, 19] solvent model during the docking procedure which faster

The protein-ligand binding energy ΔG in kcal/mol,
calculated by equation (1) in MMFF94 force field in vacuum and for the two solvent
models SGB [18, 19] and PCM [20, 21] for the different techniques of hydrogen atoms
incorporation into protein structures

PDB id	APLITE			MModel		
	vacuum	SGB	PCM	vacuum	SGB	PCM
1C5Y	−113,51	−45,4	−43,33	−152,94	−73,9	−70,82
1DWC	−57,36	1,78	2,78	−68,2	−2,41	−0,26
1F5L	−98,6	−50,5	−50,15	−119,56	−63,45	−62,29
1O3P	−114,3	−41,46	−38,4	−148,42	−72,02	−71,14
1SQO	−98,26	−45,18	−42,8	−115	−58,47	−56,06
1TOM	−147,06	−43,79	−40,54	−162,78	−49,13	−44,45
1VJ9	−133,15	−33,03	−29,49	−166,25	−53,35	−48,93
1VJA	−134,39	−46,73	−42,52	−166	−67,71	−61,58
2P94	−43,11	−14,57	−10,84	−46,74	−8,98	−6,88
3CEN	−49,74	−18,75	−15,51	−51,01	−14,15	−10,24
4FSW	−43,81	−3,15	2,71	−41,97	−2,96	1,64
4FT0	103,5	11,2	16,05	104,14	13,63	19,46
4FT9	−47,87	−14,44	−8,79	−45,57	−13,33	−8,07
4FTA	89,45	0,92	5,46	88,76	−0,17	6,74
4FV5	−60,89	−10,06	−2,16	−65,08	−10,28	−0,92
4FV6	−71,81	−10,45	−3,4	−77	−13,34	−4,17

than PCM [20, 21], but less accurate. Although the solvent consideration decreases notably the binding energy differences for the different hydrogen incorporation methods (see Table 3), though the binding energy can still vary from a few kcal/mol up to 10 − 20 kcal/mol.

Taking into account all these results it can be assumed that docking should also depend on the hydrogen atoms incorporation method. To check this assumption the docking procedure is carried out for some test proteins prepared by different programs. The docking of native ligands into corresponding proteins is done by the SOL docking program using the rigid protein approach in the frame of MMFF94 force field [4–6]. The docking parameters were the same for all considered complexes. The obtained results are presented in Table 4.

Results presented in Table 4 data demonstrate that docking scores and docked ligand RMSD from its crystalline position depend on the method used for hydrogen atoms incorporation into protein structures. For example, complex 3CEN, prepared by APLITE, has a "good" RMSD value (1,55 Å) and a "good" score value (−6,12 kcal/mol), but the same complex, prepared by the AutodockTools program, has a "bad" RMSD value (4,11 Å) and a "bad" score value (−3,89 kcal/mol). So, our assumption is confirmed: docking results depend strongly on the method of hydrogen atoms incorporation into the protein PDB [2, 3] structure. It is possible that the difference of these results will be smaller if hydrogen atoms move during the docking procedure. However, realization of mobility of the target protein atoms in the docking procedure needs adoption the new docking

SOL scores and RMSD of docked ligand positions from the crystalline one after the docking procedure for four test protein-ligand complexes. Results for all hydrogen atoms incorporation methods are presented. Numbers of ligand inner torsions are adduced together with PDB ID in the brackets

PDB id	Docking results for the different programs of hydrogen atoms addition to proteins					
1C5Y (1)	APL*	ADT	Chi1	Chi2	Rdc	Mop
RMSD, Å	2,63	1,94	1,93	2,59	1,95	2,59
score, kcal/mol	−5,89	−6,43	−6,46	−7,22	−6,72	−7,04
1SQO (3)	APL*	ADT	Chi1	Chi2	Rdc	Mop
RMSD, Å	1,06	1,09	1,3	1,09	1,1	0,95
score, kcal/mol	−6,52	−7,65	−6,7	−7,95	−7,77	−7,16
1VJ9 (17)	APL*	ADT	Chi1	Chi2	Rdc	Mop
RMSD, Å	3,1	3,59	3,21	1,52	5,49	3,34
score, kcal/mol	−4,71	−3,74	−5,14	−4,53	−4,59	−4,72
3CEN (7)	APL*	ADT	Chi1	Chi2	Rdc	Mop
RMSD, Å	1,55	4,12	1,51	1,38	1,37	1,37
score, kcal/mol	−6,13	−3,9	−5,8	−5,97	−6,1	−6,08

How to read the short program names, shown in the Table 1.

approach: instead of the traditional docking technique using the preliminary calculated energy grid of probe ligand atoms in the field of the target protein (see, for example, SOL [4–6]) the straightforward docking procedure must be used. In this docking procedure the protein-ligand energy is calculated directly in the frame of the given force field for any ligand pose in the target protein active site (see, for example FLM [7,8] and SOL-T [22]). But such approach demands much more computational expenses, especially when the target protein atoms are moveable.

Conclusions

Influence of methods of hydrogen atoms incorporation into PDB [2,3] protein structures on the protein-ligand binding energy are considered in details. Six different hydrogen atoms incorporation methods are compared for a set of 16 protein-ligand complexes.

It is revealed that APLITE, MacroModel, Chimera 2 and MOPAC programs create identical protonation of amino acids residues in most cases and charge states of respective amino acid residues turn out to be the same. In contrast, Chimera 1 and Autodock make protonation completely differently and each of these two programs makes it in its own way. The Reduce program makes the same protonation as programs of the first group make for approximately half of the tested complexes.

The results show that the method of hydrogen atoms incorporation into the target protein structure taken from the Protein Data Bank affect considerably the results of molecular modelling and, in particular, the calculated protein-ligand binding energy. The differences in binding energies can be several dozen of kcal/mol, even if the protonation (created by different methods) of protein amino acid residues are identical, i.e. the same amino acid residues have the same charges. Differences in positions of hydrogen atoms added to the protein by different methods are significant. RMSD for all protein hydrogen atoms is about of 0,7 Å. Moreover, a large part (over a quarter) of all hydrogen atoms differ strongly in their positions (1,1 − 1,8 Å).

Taking into account the mobility of protein hydrogen atoms can lead to a significant change in the binding energy of about dozens of kcal/mol for some complexes and it does not reduce a strong influence of hydrogen atoms incorporation method on the calculated protein-ligand binding energy.

In addition, the protein hydrogen atoms mobility does not affect significantly the value of RMSD between the locally optimized native (crystallized) ligand position and the initial native ligand position crystallized in the complex. The influence of hydrogen atoms incorporation method on the calculated protein-ligand binding energies somewhat reduced by taking into account interaction with water solvent, but nevertheless binding energies differ up to several kcal/mol depending on the hydrogen atoms incorporation method for many tested complexes. Moreover, the difference in binding energies reaches 20 − 30 kcal/mol for certain complexes even if respective amino acid residues have the same charges but different hydrogen atoms positions due to usage of different hydrogen atoms incorporation methods. The different protein hydrogen atoms positions affect strongly the docking results: docked ligand poses as well as the estimated the protein-ligand binding energies (scores).

It is also shown that the difference in binding energies when using two implicit solvent models, SGB [18, 19] and PCM [20, 21] is only a few kcal/mol for most of the complexes. So, for the future docking programs with solvent accounting it is possible to use the SGB solvent model which is faster than the PCM one.

The discovered effects of the influence of hydrogen atoms incorporation methods should be considered preparing the three-dimensional full-atomic target protein model and carrying out the molecular modelling of the protein-ligand interaction including docking procedure.

References

1. Sadovnichii V.A., Sulimov V.B. Supercomputing Technologies in Medicine. *In Supercomputing Technologies in Science, Education, and Industry*, Moscow, Moscow University Publishing, 2009, pp. 16–23. (in Russian)
2. *Protein Data Bank.* Available at: http://www.rcsb.org/ (accessed October 27, 2016)
3. 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, pp. 235–242. DOI: 10.1093/nar/28.1.235

4. Romanov A.N., Kondakova O.A., Grigoriev F.V., Sulimov A.V., Lushekina S.V., Martynov Y.B., Sulimov V.B. The SOL Docking Package for Computer-Aided Drug Design. *Numerical Methods and Programming*, 2008, vol. 9, no. 2, pp. 64–84. (in Russian)

5. Oferkin I.V., Sulimov A.V., Kondakova O.A., Sulimov V.B. Implementation of Parallel Computing for Docking Programs SOLGRID and SOL. *Numerical Methods and Programming*, 2011, vol. 12, no. 1, pp. 205–219. (in Russian)

6. Sulimov A.V., Kutov D.C., Oferkin I.V., Katkova E.V., Sulimov V.B. Application of the Docking Program SOL for CSAR Benchmark. *Journal of Chemical Information and Modeling*, 2013, 53, pp. 1946–1956.

7. Oferkin I.V., Sulimov A.V., Katkova E.V., Kutov D.C., Grigoriev F.V., Kondakova O.A., Sulimov V.B. Supercomputer Investigation of the Protein-Ligand System Low-Energy Minima. *Biomeditsinskaya khimiya*, 2015, vol. 61, no. 6, pp. 712–716. (in Russian)

8. Oferkin I.V., Katkova E.V., Sulimov A.V., Kutov D.C., Sobolev S.I., Voevodin V.V., Sulimov V.B. Evaluation of Docking Target Functions by the Comprehensive Investigation of Protein-Ligand Energy Minima. *Advances in Bioinformatics*, 2015, vol. 2015, Article ID 126858, 12 p. DOI: 10.1155/2015/126858

9. Sastry G.M., Adzhigirey M., Day T., Annabhimoju R., Sherman W. Protein and Ligand Preparation: Parameters, Protocols, and Influence on Virtual Screening Enrichments. *Journal of Computer-Aided Molecular Design*, 2013, vol. 27, no. 3, pp. 221–234.

10. Pettersen E.F., Goddard T.D., Huang C.C., Couch G.S., Greenblatt D.M., Meng E.C., Ferrin T.E. UCSF Chimera – a Visualization System for Exploratory Research and Analysis. *Journal of Computational Chemistry*, 2004, vol. 25, no. 13, pp. 1605–1612.

11. Word J.M., Lovell S.C., Richardson J.S., Richardson D.C. Asparagine and Glutamine: Using Hydrogen Atom Contacts in the Choice of Side-Chain Amide Orientation. *Journal of Molecular Biology*, 1999, vol. 285, no. 4, pp. 1735–1747.

12. Morris G.M., Huey R., Lindstrom W., Sanner M.F., Belew R.K., Goodsell D.S., Olson A.J. AutoDock4 and AutoDockTools4: Automated Docking with Selective Receptor Flexibility. *Journal of Computational Chemistry*, 2009, vol. 16, pp. 2785–2791.

13. O'Boyle N.M., Banck M., James C.A., Morley C., Vandermeerschand T., Hutchison G.R. Open Babel: an Open Chemical Toolbox. *Journal of Cheminformatics*, 2011, vol. 3, article number 33, 14 p. DOI: 10.1186/1758-2946-3-33

14. Stewart J.J.P. Optimization of Parameters for Semiempirical Methods VI: More Modifications to the NDDO Approximations and Re-Optimization of Parameters. *Journal of Molecular Biology*, 2013, vol. 19, no. 1, pp. 1–32. DOI:10.1007/s00894-012-1667-x

15. *Avogadro: an Open-Source Molecular Builder And Visualization Tool. Version 1.1.1.* Available at: http://avogadro.openmolecules.net/ (accessed October 27, 2016).

16. Hanwell M.D., Curtis D.E., Lonie D.C., Vandermeersch T., Zurek E., Hutchison G.R. Avogadro: An Advanced Semantic Chemical Editor, Visualization, and Analysis Platform. *Journal of Cheminformatics*, 2012, vol. 4, 17 p.

17. Halgren T.A. Merck Molecular Force Field. I. Basis, Form, Scope, Parameterization, and Performance of MMFF94. *Journal of Computational Chemistry*, 1996, vol. 17, no. 5-6, pp. 490–641.

18. Romanov A.N., Jabin S.N., Martynov Y.B., Sulimov A.V., Grigoriev F.V., Sulimov V.B. Surface Generalized Bornmethod: a Simple, Fast and Precise Implicit Solvent Model Beyond the Coulomb Approximation. *The Journal of Physical Chemistry A*, 2004, vol. 108, pp. 9323–9327.

19. Kupervasser O.Yu., Zhabin S.N., Martynov Y.B., Fedulov C.M., Oferkin I.V., Sulimov A.V., Sulimov V.B. Continual Model of Solvent: the DISOLV Software-Algorithms, Implementation and Validation. *Numerical Methods and Programming*, 2011, vol. 12, pp. 246–261. (in Russian)

20. Mikhalev A.Y., Oferkin I.V., Oseledets I.V., Sulimov A.V., Tyrtyshnikov E.E., Sulimov V.B. Application of Themulticharge Approximation for Large Dense Matrices in the Framework of the Polarized Continuum Solvent Model. *Numerical Methods and Programming*, 2014, vol. 15, pp. 9–21. (in Russian)

21. Sulimov V.B., Mikhalev A.Yu., Oferkin I.V., Oseledets I.V., Sulimov A.V., Kutov D.C., Katkova E.V., Tyrtyshnikov E.E. Polarized Continuum Solvent Model: Considerable Acceleration with the Multicharge Matrix Approximation. *International Journal of Applied Engineering Research*, 2015, vol. 10, no. 24, pp. 44815–44830.

22. Oferkin I.V., Zheltkov D.A., Tyrtyshnikov E.E., Sulimov A.V., Kutov D.C., Sulimov V.B. *Bulletin of the South Ural State University. Series: Mathematical Modelling, Programming and Computer Software*, 2015, vol. 8, no. 4, pp. 83–99. DOI: 10.14529/mmp150407

УДК 004.94+539.6   DOI: 10.14529/mmp170308

О ВЛИЯНИИ СПОСОБА ПРИГОТОВЛЕНИЯ БЕЛКА-МИШЕНИ НА СВЯЗЫВАНИЕ С НИМ ЛИГАНДА

Д.К. Кутов, Е.В. Каткова, А.В. Сулимов, О.А. Кондакова, В.Б. Сулимов
ООО «Димонта», г. Москва
Научно-исследовательский вычислительный центр, Московский государственный университет имени М.В Ломоносова, г. Москва

Выбор метода подготовки белка-мишени, в частности, метод протонирования белка, может значительно повлиять как на пространственное расположение добавленных атомов водорода, так и на зарядовые состояния отдельных молекулярных групп аминокислотных остатков. Это означает, что при вычислении энергии связывания белка и лиганда различные способы подготовки белка-мишени могут привести к существенно разным значениям энергий связывания, а в случае докинга (позиционирования лиганда в активном центре белка) – к различным положениям лиганда в активном центре. В данной работе было проведено исследование влияния способа расстановки атомов водорода в белке-мишени на энергию связывания белок-лиганд. При этом исследовались случаи, когда все атомы водорода белка-мишени были зафиксированы или подвижны. В результате для тестового набора белков-мишеней, приготовленных шестью различными программами, показано, что энергия связывания белок-лиганд может существенно зависеть от способа добавления атомов водорода, и различия могут достигать 100 ккал/моль. Показано также, что учет растворителя в одной из двух континуальных моделей несколько сглаживает эти различия, но они все равно могут оказаться порядка $10-20$ ккал/моль. Более того, мы показали, что докинг лиганда, проведенный с помощь программы SOL в закристаллизованный с ним белок, приготовленный разными способами, может дать существенно различные результаты как по позиционированию лиганда, так и по оценке его энергии связывания с белком в зависимости от способа добавления атомов водорода к белку.

Ключевые слова: белок-мишень; кристаллическая структура; протонирование; энергия связывания белок-лиганд; докинг; силовое поле.

Данил Константинович Кутов, ООО «Димонта»; Научно-исследовательский вычислительный центр, Московский государственный университет имени М.В. Ломоносова (г. Москва, Российская Федерация), dk@dimonta.com.

Екатерина Владимировна Каткова, ООО «Димонта»; Научно-исследовательский вычислительный центр, Московский государственный университет имени М.В. Ломоносова (г. Москва, Российская Федерация), katkova@dimonta.com.

Алексей Владимирович Сулимов, ООО «Димонта»; Научно-исследовательский вычислительный центр, Московский государственный университет имени М.В. Ломоносова (г. Москва, Российская Федерация), as@dimonta.com.

Ольга Анатольевна Кондакова, ООО «Димонта»; Научно-исследовательский вычислительный центр, Московский государственный университет имени М.В. Ломоносова (г. Москва, Российская Федерация), ok@dimonta.com.

Владимир Борисович Сулимов, доктор физико-математических наук, ООО «Димонта»; Научно-исследовательский вычислительный центр, Московский государственный университет имени М.В. Ломоносова (г. Москва, Российская Федерация), vs@dimonta.com.

**Вестник ЮУрГУ. Серия «Математическое моделирование
и программирование» (Вестник ЮУрГУ ММП). 2017. Т. 10, № 3. С. 94–107**

107