REPORT ON THE AVAILABILITY OF QUANTUM AND GLIDE SOFTWARE ON DOCKING LIGANDS-PROTEIN

Nguyen Hoa Mi, Dang Ung Van, Le Kim Long, Lam Ngoc Thiem, Do Quang Huy

Center for Computational Chemistry, Vietnam National University, Hanoi, Vietnam

Received 17 March 2010

Abstract

A series of the complexes of human CD38's wild type, E226 and E146 mutants as well have been simulated. The biosoftwares well simulate the penetration of nicotinamide-adenine-dinucleotide (NAD) into the active site. The breaking down hydrogen bond between 2'-3' OH ribosyl and the residues replaced Glu^{226} makes NAD to be less constrained in active site and nicotinamide (NA) becomes more difficult to be cleaved and eliminates the mutant catalytic activities. The large majority of the substrate NAD is hydrolyzed to ADPR while the conversion of NAD to cADPR is not the dominant reaction catalyzed by wild-type human CD38. These results are in good agreement with the previous crystallographic analysis and the experiments quantified the catalytic activities of human CD38 and its mutants.

1. INTRODUCTION

A fundamental postulate in the classical drug design paradigm is that the effect of a drug in the human body is a consequence of the molecular recognition between a ligand (the drug) and a macromolecule (the target). The pharmacological activity of the ligand at its site of action is ultimately due to the spatial arrangement and electronic nature of its atoms, and the way these atoms interact with the biological counterpart [1]. Computational chemistry tools allow one to characterize the structure, dynamics, and energetic of the interactions between the ligand and a macromolecule as protein and DNA. For instance, molecular mechanics (MM)-based approaches can efficiently assist the discovery of new drug candidates, and these computationally inexpensive methods are nowadays routinely used in drug design [2]. However, if a description of the electronic properties is deemed necessary, there is no substitute for quantum mechanics (QM). Indeed, since QM based approaches also account for quantum electronic effects, they describe bonds forming/breaking, polarization effects, charge transfer, etc., and usually estimate molecular energies more accurately[3 - 5]. QM methods are also fundamental to studying biological reactions, as quantum electronic effects must be taken into account to properly describe the phenomena of bonds forming/breaking. An excellent overview of target-related applications of first principles quantum chemical methods in drug design is presented in [6].

By various commercial and/or academical software's combining QM and MM based tools not only the position and pose of ligands binding on protein but also the inhibition constant as IC50 could be predicted in a rather agreement with spectroscopy experiments where $IC50 = 10^{E\_bind}/5.85$, E_{bind} is binding free energy between ligand and protein. These softwares make ability to do predictive computations of complicated biochemical processes. In this study, we employed computational method softwares, namely GLIDE [7], QUANTUM [8], WHAT IF [9], SWISS-PDB Viewer [10] and HYPERCHEM [11] to simulate the structure and energy of the enzymatic domain of human CD38 complexed with relevant ligands related to its multitude of catalytic activates. Part II introduces briefly the object ligand - CD38 complexes. Fortunately, there are series of experimental study on structure and activity of CD38 and its mutants are publicized and atomic coordinates and structure factors have been deposited at the Protein Data Bank (PDB). These data and a large amount of other ligand-protein complexes deposited in PDB have been used to assess the computational procedure in Part III. The calculated results are presented in Part IV including the cADPR hydrolase, NADase (NAD glycohydrolase) – the intermediate Michaelis complex, the activation of the intermediate Michaelis complex and E146 mutants' cyclase and hydrolyse activity. In conclusion, some limitation of the computational procedure is discussed to suggest a demand for increasing the prediction ability of the softwares.

2. LIGAND-PROTEIN COMPLEXES

CD38 was described first as an antigen that is involved in a host of lymphocyte functions including

differentiation, proliferation, and apoptosis [12]. Its expression has since been found to be widespread among nonhematopoietic tissues as well [13]. In addition to the antigenic functions, CD38 also possesses a multitude of enzymatic activities [14]. It catalyzes not only the hydrolysis of NAD[i] and cADPR to ADPR, but also the cyclization of NAD a long linear molecule, and its analog, NGD, to produce a compact cyclic nucleotide, cADPR and cGDPR, respectively. CD38 has also base-exchange activity that is responsible for synthesizing NAADP from NADP in acidic pH [14].

The active site of human CD38 has been biochemically and structurally characterized [15, 16]. GLU^{226} is identified as the catalytic residue, because its mutation to other residues essentially eliminates all its catalytic activities [15]. SER^{193} is also important for catalysis as its mutation to alanine also greatly reduces enzyme activities [17]. Conversion of NAD to cADPR, however, is not the dominant reaction catalyzed by wild-type human CD38. NAD will be conversed to cADPR, however is not the dominant reaction catalyzed by wild-type human CD38. In fact, the large majority of the substrate NAD is hydrolyzed to ADPR [18]. Completely the opposite is observed when NGD, an analog of NAD, is used as substrate. The dominant reaction is now cyclization instead of hydrolysis, producing cGDPR as the major product [19]. Considering the similarity of NGD and NAD, which differ only in the purine rings, it is puzzling why the reactions are so different [20]. GLU^{146} is a conserved residue present in the active site of CD38. Its replacement with phenylalanine greatly enhanced the cyclization activity to a level similar to that of the NAD hydrolysis activity. A series of additional replacements was made at the Glu-146 position including alanine (E146A), asparagine (E146N), glycine (E146G), aspartic acid (E146D), phenylalanine (E146F) and leucine (E146L) [21]. All the mutants exhibited enhanced cyclase activity to various degrees, whereas the hydrolysis activity was inhibited greatly. E146A showed the highest cyclase activity, which was more than 3-fold higher than its hydrolysis activity. All mutants also cyclized NGD to produce cGDPR. This activity was enhanced likewise, with E146A showing more than 9-fold higher activity than the wild type. In addition to NAD, CD38 also hydrolyzed cADPR effectively, and this activity was correspondingly depressed in the mutants. When all the mutants were considered, the two cyclase activities and the two hydrolase activities were correlated linearly. The Glu-146 replacements, however, only minimally affected the base-exchange activity that is responsible for synthesizing NAADP[21]. Unfortunately, E146-mutant's structure has not yet been deposited in Protein Data Bank. Homology modeling was used to assess possible structural changes at the active site of E146A[21].

In this study, we employed the structures of the enzymatic domain of human CD38's wild-type and its E226 mutants complexed with the relevant ligands, that is NAD, ADPR, cADPR, NGD, GDPR, cGDPR, EPE, NMN and N1C by x-ray crystallography [15 - 21, 23, 25]. The complex of E226G - a mutant of CD38 received by replacement of Glu^{226} by glycine and NMN - a substrate of CD38 (code in PDB is 2HCT), the complex of E226Q mutant of CD38 and cADPR (code 2O3Q), the complex E226D-cADPR (code 2O3R), E226G-cADPR (code 2O3S) and other ligand-protein complexes as well deposited on Protein Data Bank have been used to verify the computational procedure. The mutants investigated computationally in this study are E146A, E146N, E146G, E146L, E146D, E146F and E146K, E146Q. Two latters were obtained by replacing Glu^{146} by lysine (K) and glutamine (Q), respectively. These complexes together with the above mentioned complexes provided a step-by-step description of the catalytic processes involved in the synthesis and hydrolysis of cADPR.

3. COMPUTATIONAL PROCEDURES

The calculation procedure includes three core algorithm: i) the replacement of each residue in active sites by other one then make a geometrical optimiziation which simulates the site-directed mutagenesis technique and ii) docking ligand on mutants to determine the docking poses and iii) calculating the binding energy of the obtained complexes. Fortunately, all three algorithms could be received on web in the form of source code, executive file and/or online calculation. Depending on the concrete algorithm the results received by these softwares may be different. Though there are many articles presented the studies on the reliability of various bio –chemistry softwares applied to a

[i] The abbreviation used are: NA, nicotinamide, NAD, nicotinamide-adenine-dinucleotide, NADP, nicotinamide adenine dinucleotide phosphate, NAADP, nicotinic acid adenine dinucleotide phosphate; G1R, GDPR, guanosine diphosphoribose, APR, ADPR, adenosine-5-diphosphoribose, ADPRI, adenosine-5-diphosphoribose intermediate, EPE 4-(2-hydroxyethyl)-1-piperazine ethanesulfonic acid, NGD nicotinamide guanine dinucleotide, NMN nicotinamide mononucleotide, CXR, cADPR, cyclic adenosine diphosphate-ribose, CGR, cGDPR cyclic guanosine diphosphate-ribose, N1C, N1-cyclic inosine 5'-diphosphoribose, 1yh3, CD38'wild-type.

large amount of proteins of different kinds and shown the ability of each software, this article pays attention to the software's reliability applied to a narrow branch of proteins including the complexes between various ligands and CD38 and its mutants as well. Relating to the computational characterization of the CD38's multitude of catalytic activites, we should choose the softwares could give well prediction of the mutant structure based on native protein structure – the site-directed mutagenesis - and of the docking pose of ligand on active sites.

Docking - The ligand docking on CD38 and its mutants is predicted on QUANTUM 3.3 (DEMO version)[8] and GLIDE[7] version 4.51.08 as well. There are little differences between two softwares in the preparation of the proteins and ligands. In both cases we applied the rigid protein model and follow the docking calculation procedure to receive model structure of the ligand-protein complex. The mutant–ligand complexes as: 2hct, 2o3q, 2o3r and 2o3s, 2o3t, 2o3u, 2i65, 2pgl and wild-type CD38 –ligand complexes as: 2pgj, 2ef1, 2i66 and 2i67 deposited in PDB were chosen to be testing samples of software reliability.

As the workspace structure consists of a receptor only, there is no default center for the enclosing box. The box will not be displayed until you have specified a grid center by selecting residues or proposed ligand position. Surprisingly, the docking results depend essentially on the position of grid center, especially in QUANTUM calculation of long chain and flexible ligands. In the cases we are not sure of the proposed ligand position we may select the center of the grid box by selecting any atom that lies approximately in the middle of the active site and all active site atoms that are on the surface of the protein should be covered by a grid box, or at least all important chemical groups of the active site should lie inside the grid box. Perhaps a micro genetic algorithm loop should be used to find the grid center giving the pose of maximum binding free energy.

Among the output data of QUANTUM we can find IC50 ((Mol/L), Ebind (kJ/mol) – the binding free energy including Ees (kJ/mol) – the electrostatic and solvation energy, Evdw (kJ/mol) – the short-range electrostatic and exchange and Van der Waals energies, TdS (kJ/mol) – the entropy contribution, Etor (kJ/mol) – the ligand internal energy change. We received also the total charge Q, mass M, number of flexible bonds of the ligand and RMSD (A) - the root mean square distance between the initial and final position. In another way, GLIDE gives GLIDE score includes standard precision (SP) and extra precision (XP). Glides core is given by:

$$\text{Score} = a * \text{vdW} + b * \text{Coul} + \text{Lipo} + \text{Hbond} + \text{Metal} + \text{Rewards} + \text{RotB} + \text{Site},$$

where vdW is van der Waals interaction energy, Coul is Coulomb interaction energy, Lipo is lipophilic-contact plus phobic-attractive term, HBond is hydrogen-bonding term, Metal is metal-binding term (usually a reward), Rewards is various reward or penalty terms, RotB is penalty for freezing rotatable bonds, Site is polar interactions in the active site and a = 0.063, b = 0.120 for Standard Precision (SP) Glide 4.5.

In order to compare the reliability of the softwares we used RMSD (Å) - the root mean square distance between the solved position deposited in PDB and the model position given by docking software. The calculation results are presented in table 2. As most docking softwares gives some predicted docking sites of the ligand Table 2 presented the RMSD and score of two best ones. It is clearly that GLIDE gives excellent results for cycle ligands and in most cases gives RMSD smaller than QUANTUM. In addition, both softwares give RMSD > 2.0 Å for the complexes of CD38-wild type, especially, with NGD (code 2i66) where the active sites of both molecules were saturated with substrate NGD+, and reaction proceeded in the crystal. So that molecule B contains two nucleotides, a GDP-ribose intermediate and a hydrolyzed product, GDPR, whereas molecule A contains GDPR dimer[17]. The docking calculation of only one GDPR molecule would never give good agreement with crystallographic data.

It should be noted that in some cases (the underline numbers in table 1) the pose of smaller deviation has lower score (GLIDE) or higher free energy (QUANTUM). Fig. 1 displays, for example, two highest score docking poses of cADPR on E226Q mutant obtained by QUANTUM. In most of these cases the docking pose pairs of small binding free energy difference (~2 KJ/mol for QUANTUM) or small score difference (~0.5 for GLIDE). It can be regarded approximately as the *indefiniteness* of the docking data given by the software in the cases of very large and extremely flexible ligands. Therefore, using QUANTUM and GLIDE as well to predict the protein-ligand complex structures, it should be taken care the docking site pairs of small binding free energy difference (QUANTUM) or small score difference (GLIDE).

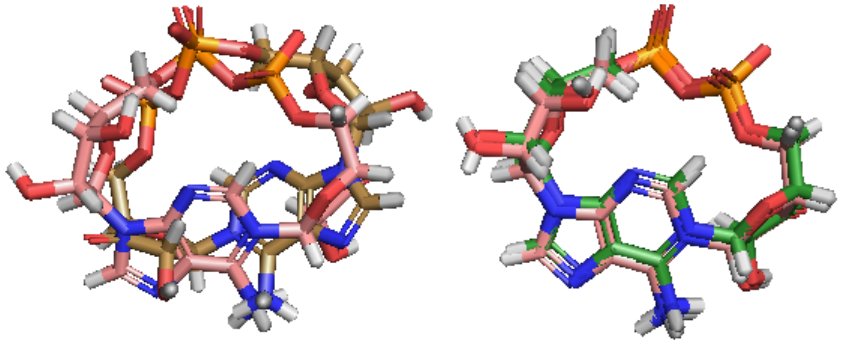
4. DISCUSSION

In order to continue the prediction study at quantitative degrees, we should deal with the potential energy surface of ligands in active site in

which the most difficult work would be finding the transition state. For locating a transition state

Table 1: Prediction results of docking ligands on CD38 and its mutants

Protein	Mutant	Ligand	QUANTUM		GLIDE (SP)	
			RMSD^{+}	E_{bind}	RMSD^{++}	Gscore
2o3r	E226D	CXR	0.3198	-37.0485	0.2662	-8.75
			6.1247	-34.4071	2.3827	-8.60
2o3s	E226G	CXR	6.1885	-37.5786	0.2486	-9.55
			0.6142	-36.7304	0.5771	-9.43
2o3t	E226Q	CGR	7.4470	-34.2169	0.1589	-14.99
			0.6605	-33.9033	0.3874	-12.78
2pgj	1yh3	N1C	1.2325	-34.6732	0.2251	-9.59
			8.1953	-30.8546	8.0911	-4.13
2pgl	E226Q	N1C	1.2910	-35.0580	0.2088	-11.76
			8.6005	-30.9110	5.6327	-2.94
2o3q	E226Q	CXR	6.0261	-36.5327	1.3044	-6.22
			0.3264	-34.6329	0.5761	-5.49
2hct	E226G	NMN	0.6842	-43.2520	0.7361	-11.40
			5.3212	-32.1916	1.0533	-11.40
2o3u	E226Q	NGD	1.3438	-50.9927	3.0937	-10.15
			4.7932	-49.6694	3.7598	-9.95
2ef1	1yh3	EPE	1.2447	-22.6017	5.2035	-3.81
			3.6816	-20.2168	4.8656	-3.69
2i65	E226Q	NAD	0.9585	-45.9585		
			2.4248	-45.5675		
2i66	1yh3	G1R	1.2910	-35.0580	4.9718	-9.04
			8.6005	-30.9110	2.5998	-8.96
2i67	1yh3	APR	3.2385	-38.1995	2.3588	-7.66
			6.4285	-37.9830	2.1810	-7.59



a b

Figure 2. Comparison of cADPR-ligand in two configuration of largest binding free energy docked on E226Q mutant by QUANTUM and the poses deposited on Protein Data Bank (code 2o3q). Being of smaller binding free energy (-34.6329 KJ/mol) b-configuration is in a much better agreement with PDB data than a-

+ Comparing with the prepared ligand position
++ Comparing with the initial ligand position

structure the structures of the react, product and even transition state guess [15] should be given using linear and/or quadratic *synchronous transit* [16] algorithm. Without giving the react and product structures locating transition states is somewhat more complex. Guessing a good trial structure is the hardest part of the work. The Hessian should be computed at best guess as to what the transition state (T.S.) should be. This lets you verify you have guessed a structure with one and only one negative

curvature. An initial guess at a transition state structure might in the simplest case be based on chemical intuition; that is, what one already knows about the reaction mechanism. It seems to exceed the current biosoftwares' capability and also a demand suggested to the softwares builders though the modern quantum chemistry softwares could be applied successfully to calculate PES of the small molecule system. The second demand is an entire flexible ligand docking techniques which could give ability to realize a constrained docking in which we can keep some atomic distance constant.

Acknowledgments: *We thank Qun Liu (Cornell University) et al. for giving CD38's mutants crystallographic data and their related articles, WHAT IF's authorities for on-line calculation, QUANTUM PHARMACEUTICALS for supplying QUANTUM 3.3.0 demo version, Schrodinger for evaluation license of MAESTRO, GLIDE, JAGUAR, LIGPREP.*

REFERENCES

1. Bohm, H. J.; Klebe, G. Angew. Chem., Int. Ed., 35, 2589 (1996).
2. W. L. Jorgensen. Science, 303, 1813 (2004).
3. Sherwood, P. Methods and algorithms of quantum chemistry; John von Neumann Institute for Computing: Juelich, 2000; Vol. 1.
4. W. L. Jorgensen, and J. Tirado-Rives. J. Am. Chem. Soc., 110, 1657 - 1666 (1988).
5. B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Kaplus. J. Comput. Chem., 4, 187-21 (1983).
6. A. Cavalli, P, Carloni, and Maurizio Recanatini. Chemical Reviews, 106, 3497 - 3519 (2006).
7. GLIDE: Schrodinger, LLC: Portland, OR, (2000).
8. QUANTUM 3.3 Docking/Library Screening Software Quantum Pharmaceuticals 2007.
9. WHAT IF http://swift.cmbi.ru.nl/gv/ .
10. SWISS-PdbViewer.
11. HYPERCHEM^{TM} Release 7.0 for Windows Molecular Modeling System Copyright @2002 Hypercube,Inc.
12. F. Malavasi, A. Funaro, S. Roggero, A. Horenstein, L. Calosso, and K. Mehta. Immunol. Today, **15**, 95 - 97 (1994).
13. K. M. Khoo, and C. F. Chang. Brain Res., 821, 17 - 25 (1999).
14. H. C. Lee. Mol. Med., 12 (11-12) 317 - 323 (2006).
15. C. Munshi, R. Aarhus, R. Graeff, T. F. Walseth, D. Levitt, and H. C. Lee. J. Biol. Chem., 275, 21566 - 21571 (2000).
16. Q. Liu, I. A. Kriksunov, R. Graeff, C. Munshi, H. C. Lee, and Q. Hao. Structure, 13, 1331 - 1339 (2005).
17. Q. Liu, I. A. Kriksunov, R. Graeff, H. C. Lee, and Q. Hao. J. Biol. Chem., 281, 32861 - 32869 (2006).
18. M. Howard, J. C. Grimaldi, J. F. Bazan, F. E. Lund, L. Santos- Argumendo, R. M. E. Parkhouse, T. F. Walseth, and H. C. Lee. Science, 262, 1056 - 1059 (1993).
19. R. Graeff, T. F. Walseth, K. Fryxell, W. D. Branton, and H. C. Lee. J. Biol. Chem., 269, 30260 - 30267 (1994).
20. Q. Liu, I. A. Kriksunov, R. Graeff, H. C. Lee, and Q. Hao. J. Biol. Chem., 282, 5853 - 5861 (2007).
21. R. M. Graeff, C. Munshi, R. Aarhus, M. Johns, and H. C. Lee. J. Biol. Chem. (2001) 276, 12169-12173 (2001).
22. C. Peng, H. B. Schlegel. Israel J. Chem., **33**, 449 (1993).
23. T. A. Halgren, W. N. Lipscomb. Chem. Phys. Lett., (1977) 225 (1977).

Corresponding author: **Nguyen Hoa Mi**
Center for Computational Chemistry, Vietnam National University, Hanoi, Vietnam
19 Le Thanh Tong, Hoan Kiem District, Hanoi Vietnam
Email: minguyenhoa@yahoo.com.vn