QM Scoring Function and its Application to TOP1 Inhibitors

Ky-Youb Nam, Inhee Choi, and Kwang-Hwi Cho[†,*]

Drug Discovery Division, Bioinformatics & Molecular Design Research Center, Seoul 120-749, Korea
[†]Department of Bioinformatics and Life Science and CAMD Research Center, Soongsil University, Seoul 156-743, Korea
[*]E-mail: chokh@ssu.ac.kr
Received December 9, 2010, Accepted January 24, 2011

Because of the dramatic increase in the number of therapeutic target molecules and the need to study their 3D structures, computational screening methods have gained crucial importance for the hit identification of drug discovery.[1,2] However, one of the obstacles in virtual screening (VS) is the low correlation between the biological activities and the docking scores. Although VS techniques continue to be developed and improved with respect to the search methods and score functions,[3] most current empirical scores such as *Ludi*[4] and *PLP*[5] are still ineffective for describing electron transfer phenomena such as *cation-π* interactions. Consequently, quantum mechanical(QM) methods have been applied to protein-ligand complex.[6,7] However, due to their sizes of systems and amount of computational resources needed, applying QM method to biological systems is still limited. To overcome these problems, this study suggests a suitable QM based scoring strategy and applies this strategy to the topoisomerase I (TOP1)-inhibitor complex. TOP1 is an essential enzyme that relaxes DNA supercoiling and relieves torsional strain during DNA processes.[8,9] TOP1-inhibitor interactions are very good examples for validating the QM scoring strategy because the interaction involves three different molecules, protein, DNA, and ligand, thus making it difficult to describe by empirical score functions. TOP1 inhibitors stabilize the binary complex of TOP1 and DNA by intercalating with the DNA base pairs and by interacting with the residues of TOP1. The X-ray crystal structures of a human TOP1–DNA complex bound with 3 inhibitors-topotecan (L1),[10] camptothecin (L2),[11] and indolocarbazole (L3)[10]-have been previously reported with their biological activities (pK_i) (Fig. 1). The crystal structures of the TOP1-L1, L2, and L3 complexes obtained from the Protein Data Bank[12] (PDB ID: 1T8I, 1K4T, and 1SEU, respectively) were used as the initial structures. Four methods-Ludi,[4] *LIGSCORE1*, *LIGSCORE2*,[13] and *PLP*[5]- were tested and the correlation (R^2) between the calculated scores and the observed pK_i values is shown in Table 1. The best score was obtained for *PLP* with R^2 values of 0.89.

QM calculations were performed for the same complexes using the following procedure. The complexes (protein-DNA-ligand) were energy minimized by use of molecular mechanics within Discovery Studio 2.1 (Accelrys).[14] With the exception of the residues within 5 Å from the ligand,[15] other residues were eliminated and discontinuous portions

between the amino acids and the DNA were replaced with hydrogen atoms to make the complex sufficiently small for the QM treatment. Three different sets of calculations were performed for each complex: TOP1_fragment-DNA_fragment–ligand ($G^{Complex}$), TOP1_fragment-DNA_fragment (G^{TOP1}), and ligand (G^{Ligand}). Frequency calculations were introduced to calculate the entropic contribution to the total free energy. The calculations were carried out using GAUSSIAN03.[16] During the QM calculation, the coordinates of the heavy atoms of TOP1 and DNA parts were fixed whereas the other coordinates were relaxed. The total energy and interaction energy were evaluated at the HF level using the STO-3G and 4-21G basis sets. Relatively small basis sets were used because the systems are still fairly large for the application of QM.

The binding energy was calculated as follows:

$$\Delta E(bind) = E^{Complex} - (E^{TOP1} + E^{Ligand}) \qquad (1)$$

The binding energies were computed to investigated the structure activity correlations, and frequency calculations were performed to introduce entropy terms into the free energy calculation, $G = H - TS$. Entropy calculations were
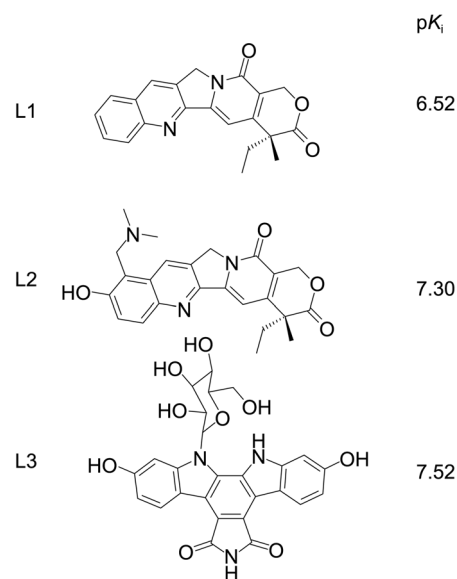
Figure 1. Structures of the three ligands studied-Topotecan (L1), camptothecin (L2), and indolocarbazole (L3)-and their experimental inhibition constants.

1128   *Bull. Korean Chem. Soc.* **2011**, Vol. 32, No. 4

Communications to the Editor

Table 1. Ludi and other docking scores of TOP1 inhibitors

	pK_i	*Ludi*	*LigScore1*	*LigScore2*	*PLP*
L1 (1T8I)	6.52	486	4.06	3.42	132.23
L2 (1K4T)	7.30	443	5.64	6.87	181.75
L3 (1SEU)	7.52	616	5.01	6.18	175.15
R^2		0.22	0.66	0.85	0.89

Table 2. Quantum mechanical binding energy score for TOP1 inhibitors (energies are in kcal/mol)

		STO-3G		4-21G	
	pK_i	HF^a	*Free Energy*^{b}	HF	*Free Energy*
L1(1T8I)	6.52	−23.75	−3.15	−43.19	−21.28
L2(1K4T)	7.30	−51.74	−28.50	−81.38	−58.89
L3(1SEU)	7.52	−218.67	−36.98	−303.65	−77.99
R^2		0.584	0.999	0.587	0.984

^{a}Binding energy (HF): $\Delta E_{BasisSet}^{i} = E_{BasisSet}^{i,Complex} - (E_{BasisSet}^{TOP1} + E_{BasisSet}^{ith\ Ligand})$, $i = L1, L2, L3, BasisSet = STO - 3G, 4 - 21G$
^{b}Free energy of binding (Free Energy):
$\Delta G_{BasisSet}^{i} = G_{Vibration\ lEntropy}^{i,Complex} - (G_{Vibration\ lEntropy}^{TOP1} + E_{Vibration\ lEntropy}^{ith\ Ligand})$.
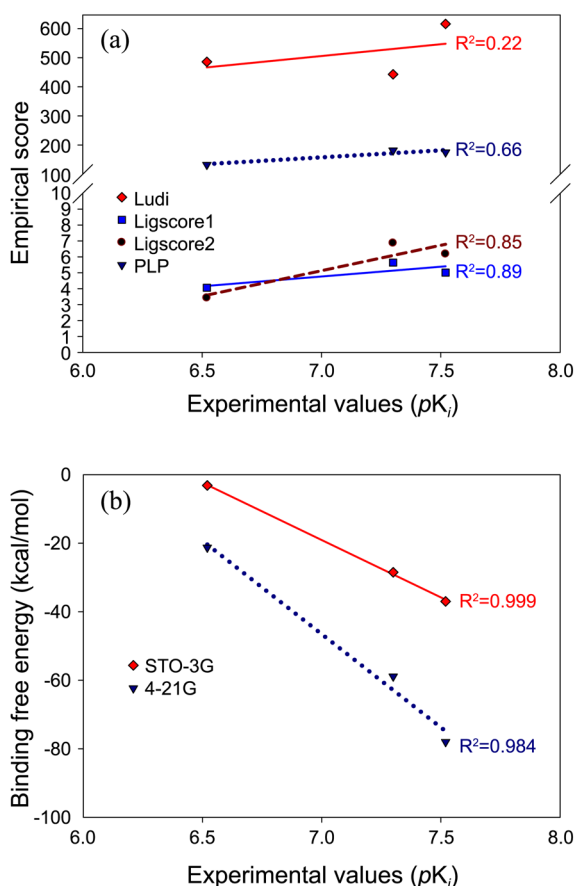
Figure 2. pK_i versus predicted scores: (a) *Ludi*, *LigScore1*, *LigScore2*, and *PLP*; (b) free energy of binding with STO-3G and 4-21G basis sets.

performed by considering the contributions from the rotational and vibrational motions. Changes in entropy due to the decreased flexibility of the ligand upon binding are commonly

regarded as a configurational entropy.[17]

The quantum mechanical binding energy and the free energy of binding are listed in Table 2. The correlation between the experimental activity and the theoretical free energy of binding was calculated and the results with the 4-21G basis set showed relatively low statistics ($R_{4\text{-}21G,HF}^{2} = 0.587$, and $R_{4\text{-}21G,FreeEnergy}^{2} = 0.984$) compared to the other basis set ($R_{STO\text{-}3G,HF}^{2} = 0.584$, and $R_{STO\text{-}3G,FreeEnergy}^{2} = 0.999$). Figure 2 shows a better comparison between the empirical and the QM-based methods.

To overcome the limitations of the empirical scoring function that is usually very difficult to correlate to biological activities, we suggest QM-based methods for describing the interactions between proteins and ligands. In this study, the QM-based method, especially with the entropic term included, shows very good correlation with experimentally determined biological activities. Once a binding mode is given by a docking program, the QM-based method can accurately estimate the interactions between the ligand and the target protein.

References

1. Douglas B. Kitchen, D. B.; Decornez, H.; Furr, J. R.; Bajorath, J. *Nat. Rev. Drug Discov.* **2004**, *11*, 935.
2. Walters, W. P.; Stahl, M. T.; Murcko, M. A. *Drug Discov. Today* **1998**, *3*, 160.
3. McInnes, C. *Curr. Opin. Chem. Biol.* **2007**, *11*, 494.
4. Böhm, H.-J. *J. Comput. Aid. Mol. Des.* **1994**, *8*, 243.
5. Gehlhaar, D. K.; Verkhivker, G. M.; Rejto, P. A.; Sherman, C. J.; Fogel, D. R.; Fogel, L. J.; Freer, S. T. *Chem. Biol.* **1995**, *2*, 317.
6. Friesner R. A.; Beachy M. D. *Curr. Opin. Struct. Biol.* **1998**, *8*, 257.
7. (a) Raha, K.; Merz, K. M., Jr. *J. Am. Chem. Soc.* **2004**, *126*, 1020. (b) Gogonea, V.; Merz, K. M., Jr. *J. Phys. Chem. A* **1999**, *103*, 5171.
8. Wang, J. C. *Annu. Rev. Biochem.* **1996**, *65*, 635.
9. Champoux, J. J. *Annu. Rev. Biochem.* **2001**, *70*, 369.
10. Staker, B. L.; Feese, M. D.; Cushman, M.; Pommier, Y.; Zembower, D.; Stewart, L.; Burgin, A. B. Jr. *J. Med. Chem.* **2005**, *48*, 2336.
11. Staker, B. L.; Hjerrild, K.; Feese, M. D.; Behnke, C. A.; Burgin, A. B., Jr.; Stewart, L. *Proc. Natl. Acad. Sci. USA* **2002**, *99*(24), 15387.
12. Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic Acids Research* **2000**, *28*, 235.
13. Discovery Studio 2.1 of Accelrys Inc., San Diego, CA, http://www.accelrys.com.
14. Venkatachalam, C. M.; Jiang, X.; Oldfield, T.; Waldman, M. *J. Mol. Graph. Model.* **2003**, *21*, 289.
15. Frisch, M. J. *et al.* GAUSSIAN 03, Revision B2; Gaussian Inc.: Pittsburgh, PA, 2003.
16. The following residues were included in the calculations: Glu356, Lys374, Lys532, Asp533, Thr718, Asn722, Tyr723, DNA Chain B-T10, Chain C-G11, Chain D-C112~A113.
17. (a) Chen, W.; Chang, C.; Gilson, M. K. *Biophys. J.* **2004**, *87*, 3035. (b) Chen, W.; Chang, C.; Gilson, M. K. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 1534.