Available online at www.sciencedirect.com
ScienceDirect
Bioorganic & Medicinal Chemistry Letters
Bioorganic & Medicinal Chemistry Letters 16 (2006) 6334–6337

Binding affinity prediction of non-peptide inhibitors of HIV-1 protease using COMBINE model introduced from peptide inhibitors

Shinya Nakamura, Isao Nakanishi* and Kazuo Kitaura

Department of Theoretical Drug Design, Graduate School of Pharmaceutical Sciences, Kyoto University, Sakyo-ku, Kyoto 606-8501, Japan

Received 19 July 2006; revised 1 September 2006; accepted 5 September 2006
Available online 5 October 2006

Abstract—Comparative binding energy (COMBINE) analysis method is one of the QSAR techniques for the prediction of biological activities of inhibitors based on interaction energies between ligands and proteins decomposed into each amino acid residue. We supposed that the predictive ability of the COMBINE method does not depend essentially on the molecular frameworks of ligands. To verify this idea, we performed the COMBINE analysis of non-peptide inhibitors of HIV-1 protease (HIVp), where the prediction model was constructed using inhibitors with a peptide scaffold as a training set. The predictive performance of the AMBER and CHARMm force fields was very high and at the same level (q^2 = 0.75, 0.67, $SDEP_{cv}$ = 0.76, 0.89, and $SDEP_{ex}$ = 0.92, 0.66, respectively). The high predictive ability of the COMBINE method for the distinct scaffold compounds is due to the informative description of the interaction energies for compounds that are located at the binding site. This result suggests that COMBINE analysis may be applied not only to the lead optimization stage but also to the lead evolution stages.

One of the purposes of drug design is to progress a seed or a lead compound to a candidate compound by improving the potency or pharmacokinetic properties, or by reducing toxicity. Intuitions or experiences of medicinal chemists have greatly contributed to this cause in many instances. On the other hand, quantitative structure–activity relationship (QSAR) analysis has been widely used to understand the relationship between biological activities and structure of compounds, and to obtain guidelines for drug design. The 3D-QSAR analysis represented by the CoMFA method[1] has become popular for ligand-based drug design since the late 1980s. Recently, structure-based drug design (SBDD), which utilizes structural information of proteins or protein–ligand complexes for drug design, is extensively used in pharmaceutical companies in conjunction with the recent progress in structural biology. Consequently, 3D-QSAR analysis has also evolved from a ligand-based method to a receptor-based one, such as the comparative binding energy analysis (COMBINE) method[2] and the method of adaptation of fields for molecular comparison (AFMoC).[3]

In the COMBINE analysis, the van der Waals and electrostatic terms of the interaction energies between a ligand and every amino acid residue of a protein are calculated. Then, the relationship between binding affinities such as K_d, K_i, or IC_{50} values and residue-based interaction energies is established by partial least squares (PLS) analysis.[4] The weights of each component of the interaction energies are treated as indexes of relative importance to describe the variation in binding affinity. It has been shown that the COMBINE analysis predicts binding free energies with good accuracy, but also identifies important amino acid residues for the improvement of affinity.[2,5–9] Thus, the COMBINE analysis is expected to provide useful information about how to improve ligand activity, in addition to predicting precise ligand affinities.

The PLS analysis implemented in the COMBINE analysis uses residue–base interaction energies between ligands and proteins. This approach is expected to predict the binding affinities of ligand molecules with various frameworks, although the application of this method has been limited to molecules with similar

Keywords: 3D-QSAR; HIV-1 protease.
* Corresponding author. Tel./fax: +81 75 753 9273; e-mail: isayan@pharm.kyoto-u.ac.jp

0960-894X/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmcl.2006.09.006

frameworks in previous studies.[2,5–9] If the interaction energies provided by the training set compounds are informative enough to describe the relationship between affinities and the energetic contributions of residues at the binding site, COMBINE analysis should predict the affinities of compounds possessing frameworks distinct from those of the training set compounds. Proteases and their inhibitors are suitable systems for the validation of predictive ability, because proteases mainly recognize substrates and inhibitors at their binding subsites such as S1, S2, S1′, S2′, etc., and the experimental affinities are accumulated. This was followed by binding affinity prediction of non-peptide inhibitors of HIV-1 protease (HIVp) based on the COMBINE analysis, where the prediction model was constructed using inhibitors with a peptide scaffold as a training set.

Gago et al. carried out the COMBINE analysis of 48 peptide inhibitors of HIVp,[5] and showed that, using a regression model generated based on a training set of 32 compounds, the binding affinities of the remaining 16 compounds could be predicted successfully, with r^2 and $SDEP_{ex}$ values of 0.90 and 0.59, respectively. Here, we used all the 48 peptide compounds as a training set, and seven non-peptide compounds with the cyclic urea and cyclic sulfonamide scaffolds were chosen[10] as a test set. The chemical structures and biological activities of the compounds used as the test set are shown in Table 1. The coordinates of the 48 HIVp complexes with peptide inhibitors were provided by Prof. F. Gago. The complex structures of three test set compounds (**1**, **2**, and **6**) with HIVp were taken from X-ray experiments (PDB codes; 1MES,[11] 1AJV,[12] and 1AJX,[12], respectively), and the others were modeled based on structurally similar complexes. The IC_{50} values of the compounds were taken from the literature.[10] Although there are several mutations in the amino acid sequences of the non-peptide complexes relative to the peptide complexes, we consistently used the same protein for the former inhibitor group as in the latter complexes. It is well-known that the non-peptide inhibitors used here interact with IleA50 and IleB50 through direct hydrogen bonds, while one water molecule mediates interaction between the protein and the peptide inhibitors (Fig. 1). Gago et al. treated this water molecule as a part of the protein. However, we incorporated this molecule into the inhibitor molecules in the following analysis, because the amide oxygen atoms of the non-peptide inhibitors fill the role of this water molecule. This treatment resulted in the interactions between the two Ile residues and the peptide compounds and the non-peptide compounds being of the same order in magnitude.

Optimization of the complex structures was performed using the Sander module in the AMBER software suite.[13] Parameter assignment was carried out according to a method reported previously;[5] briefly, the AMBER force field parameters[14] were used for the protein and water, the Generalized AMBER Force Field parameters[15] were assigned to the ligand molecules using the Antechamber module, and the partial atomic charges were derived by fitting the molecular electrostatic potential calculated by means of the AM1 method.[16] The residue-based interaction energies (van der Waals and electrostatic terms) were calculated for every ligand molecule using the Anal module. To verify whether another force field can construct a prediction model with similar

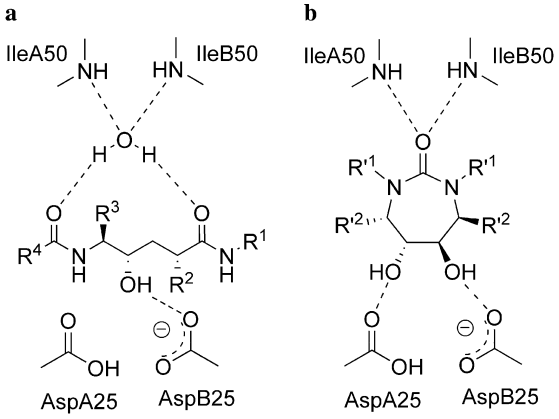
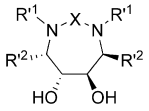


Figure 1. Hydrogen bonding arrangements between protein and inhibitor. (a) Peptide inhibitors. (b) Cyclic urea inhibitors.

Table 1. Non-peptide inhibitors used as a test set



No.	X	R′^{1}	R′^{2}	pIC_{50}[a]	pIC_{50}[b]	pIC_{50}[c]
1	CO	CH_2-4-(CH_2OH)Ph	Ph	7.82	6.72	7.04
2	CO	CH_2Ph	OPh	7.10	6.06	6.24
3	CO	CH_2Ph	CH_2Ph	5.40	5.31	5.78
4	CO	CH_2Ph	$(CH_2)_2$Ph	5.00	5.74	6.22
5	CO	CH_2-4-(CH_2OH)Ph	OPh	8.00	6.84	7.77
6	SO_2	CH_2Ph	OPh	6.70	7.02	6.84
7	SO_2	CH_2Ph	CH_2Ph	6.70	7.95	6.69

[a] Experimental value.
[b] Calculated value using AMBER force field.
[c] Calculated value using CHARMm force field.

precision, the same procedure was performed with the CHARMm force field (CHARMM.cfrc) using the CHARMm program[17] implemented in InsightII.[18] To investigate the electrostatic contribution of the solvation effect on the molecular association, the Poisson–Boltzmann equation was solved using the DelPhi program.[19] The PLS regression against these variables (residue-based interaction energies and desolvation terms) was performed using the COMBINE program.[2] Noise was reduced by zeroing the variables that had absolute values lower than 0.1 kcal mol^{-1}, and by removing any variables with a standard deviation below 0.1 kcal mol^{-1} before PLS regression.

The predictive regression models derived are shown in Tables 1 and 2, and Figure 2. As expected, these models accurately predict pIC_{50} values of the test set molecules with a scaffold differing from those used as the training set; the $SDEP_{ex}$ values are less than 1.0 and no significant outliers exist. As seen in Figure 2, the deviations of the predicted values from the experimental ones are small and the predictive performances of both the force fields are almost at the same level. This result indicates that similar functions and parameters of the van der Waals and electrostatic terms are employed in both force fields.

We think that the high predictive ability of the COMBINE method for compounds with distinct scaffolds is produced by the informative description of the interactions of the compounds at the binding site. The complex structures showed that the non-peptide inhibitors used the same binding pockets as by the peptide inhibitors; both types of compounds occupied the S1, S2, S1′, and S2′ sites and formed similar hydrogen bonds. This fact suggests that both the types of compounds interact with the protein in an analogous manner, and the training set compounds provide informative binding site features. To verify this idea, the residue-based weights of contribution for the van der Waals and electrostatic interaction energies determined by compounds restricted to the training set (predictive model) were compared with those determined by using all the 55 compounds, including the test set compounds (validation model). The results are shown in Figure 3. The contribution weights of each residue are almost the same in both the analyses; the correlation between the two is extremely high ($r^2 = 0.97$). The slight changes observed in the AspA25 and AspB25 weights have to be brought about by the differences in the hydrogen bonds with the ligands; the peptide inhibitors interact with these residues with one hydroxyl group, while the non-peptide inhibitors interact with two hydroxyl groups as shown in Figure 1. The ArgB8, AspB29, GlyB48, and GlyB49 weights changed to some extent because of the interactions with the inhibitor's substituents, which differ among the inhibitor types. These differences in residue weights between the two models are very small (0.022–0.055) and affect the predicted pIC_{50} by a value of 0.25 at most. The predictive performances of the two models are practically the same, proving that the predictive model is informative for the affinity prediction of non-peptide ligands. Thus, the application of COMBINE analysis is expected to give an accurate affinity prediction model for ligands that have distinct scaffolds, where the ligands interact with proteins at the same binding pockets as the training set molecules.

We have shown that COMBINE analysis can predict the binding affinities of test set compounds with scaffolds distinct from those of the training set compounds

Table 2. Comparison of the two regression models

Model	No. of Objects	No. of variables	No. of LVs	r^2	q^2	$SDEP_{cv}$	$SDEP_{ex}$
AMBER	48	64	2	0.82	0.75	0.76	0.92
CHARMm	47[a]	67	3	0.85	0.67	0.89	0.66

[a] The ligand missing parameter in CHARMm force field has been omitted from regression.

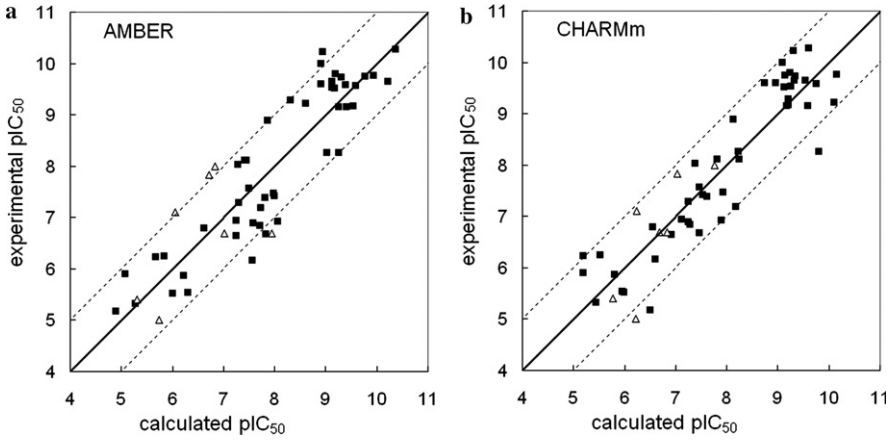


Figure 2. Correlation between experimental and calculated pIC_{50} values. Compounds in the training set and the test set are indicated by closed squares and open triangles, respectively. The dashed lines show the 10-fold error margin of the IC_{50} value. (a) AMBER force field, (b) CHARMm force field.

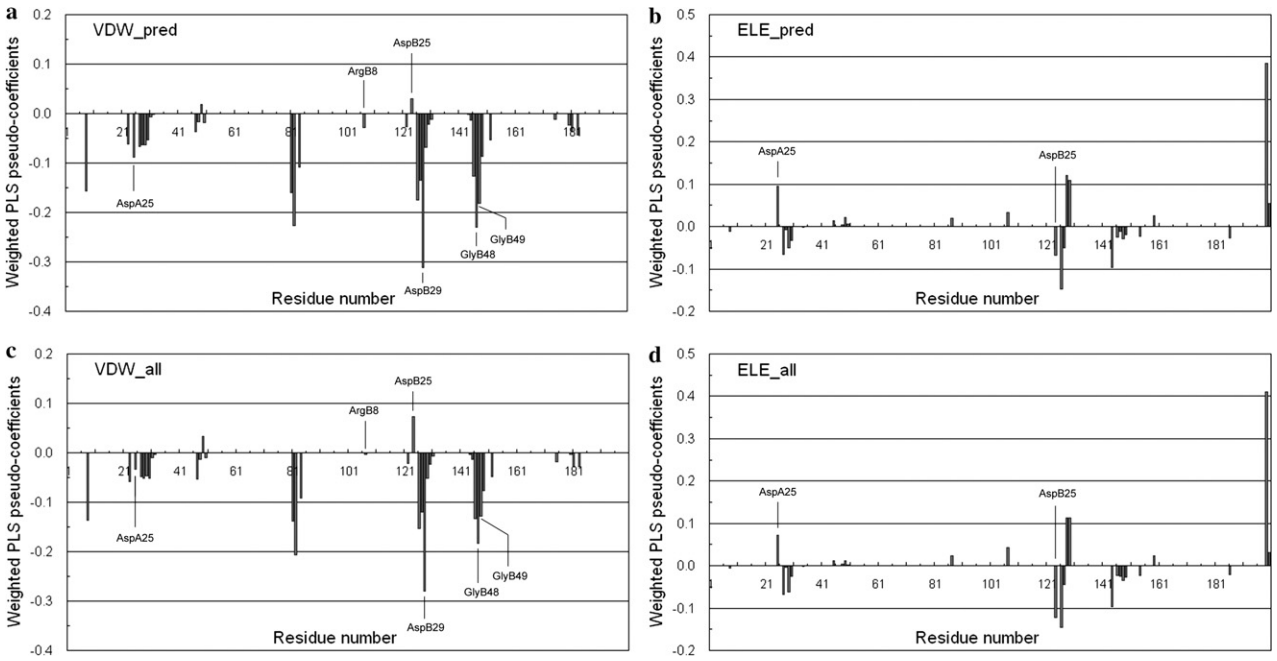


Figure 3. Comparisons of weighted PLS pseudo-coefficients. (a) Coefficients of van der Waals terms in predictive model. (b) Coefficients of electrostatic terms in predictive model. (c) Coefficients of van der Waals terms in validation model. (d) Coefficients of electrostatic terms in validation model. The electrostatic desolvation terms of ligand and protein are included in the electrostatic terms in each model as residues 199 and 200 (two from the right), respectively. Several amino acid residues with relative weight changes are labeled.

in the case of HIVp and its inhibitors. This might be an easy example, because both types of inhibitors occupy the same binding subsites where the protease recognizes the ligand molecules. We are now applying the method to other proteins that have no obvious binding subsites unlike proteases. If the binding affinities of such proteins are successfully predicted, the COMBINE analysis will contribute not only to the lead optimization stage but also to the lead evolution stages.

Acknowledgment

We are grateful to Astellas Pharma Inc. (former Fujisawa Pharmaceutical Co., Ltd.) for financial support.

References and notes

1. Cramer, R. D.; Patterson, D. E.; Bunce, J. D. *J. Am. Chem. Soc.* **1988**, *110*, 5959.
2. Ortiz, A. R.; Pisabarro, M. T.; Gago, F.; Wade, R. C. *J. Med. Chem.* **1995**, *38*, 2681.
3. Gohlke, H.; Klebe, G. *J. Med. Chem.* **2002**, *45*, 4153.
4. (a) Wold, S.; Johansson, E.; Cocchi, M. In *3D QSAR in Drug Design. Theory, Methods and Applications*; Kubinyi, H., Ed.; ESCOM Science Publishers B.V.: Leiden, 1993; pp. 523–550; (b) Cramer, R. D., III *Perspect. Drug Discovery Design* **1993**, *1*, 269.
5. Perez, C.; Pastor, M.; Ortiz, A. R.; Gago, F. *J. Med. Chem.* **1998**, *41*, 836.
6. Lozano, J. J.; Pastor, M.; Cruciani, G.; Gaedt, K.; Centeno, N. B.; Gago, F.; Sanz, F. *J. Comput. Aided. Mol. Des.* **2000**, *14*, 341.
7. Tomic, T.; Nilsson, L.; Wade, R. C. *J. Med. Chem.* **2000**, *43*, 1780.
8. Wang, T.; Wade, R. C. *J. Med. Chem.* **2001**, *44*, 961.
9. Murcia, M.; Ortiz, A. R. *J. Med. Chem.* **2004**, *47*, 805.
10. Hultén, J.; Bonham, N. M.; Nillroth, U.; Hansson, T.; Zuccarello, G.; Bouzide, A.; Åqvist, J.; Classon, B.; Danielsson, H.; Karlén, A.; Kvarnström, I.; Samuelsson, B.; Hallberg, A. *J. Med. Chem.* **1997**, *40*, 885.
11. Ala, P. J.; Huston, E. E.; Klabe, R. M.; McCabe, D. D.; Duke, J. L.; Rizzo, C. J.; Korant, B. D.; DeLoskey, R. J.; Lam, P. Y.; Hodge, C. N.; Chang, C. H. *Biochemistry* **1997**, *36*, 1573.
12. Backbro, K.; Lowgren, S.; Osterlund, K.; Atepo, J.; Unge, T.; Hulten, J.; Bonham, N. M.; Schaal, W.; Karlen, A.; Hallberg, A. *J. Med. Chem.* **1997**, *40*, 898.
13. AMBER, version 8 (UCSF); Department of Pharmaceutical Chemistry, University of California, San Francisco.
14. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179.
15. Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *J. Comput. Chem.* **2004**, *25*, 1157.
16. Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1985**, *107*, 3902.
17. Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187, Modified version in 1997.
18. InsightII, version 2000L; Accelrys Inc., San Diego, CA 92121.
19. Nicholls, A.; Honig, B. *J. Comput. Chem.* **1991**, *12*, 435.