J|A|C|S
COMMUNICATIONS

Published on Web 02/17/2004

Chemometrical Identification of Mutations in HIV-1 Reverse Transcriptase Conferring Resistance or Enhanced Sensitivity to Arylsulfonylbenzonitriles

Fátima Rodríguez-Barrios and Federico Gago*

Departamento de Farmacología, Universidad de Alcalá, E-28871 Alcalá de Henares, Madrid, Spain

Received October 6, 2003; E-mail: federico.gago@uah.es

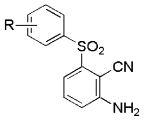
Targeting HIV-1 reverse transcriptase (RT) with nonnucleoside inhibitors (NNRTI) to combat AIDS is complicated, among other factors, by the rapid emergence of drug-resistant strains which are usually cross-resistant to other inhibitors within the class. As a consequence, the search for structurally diverse second-generation NNRTI is still being actively pursued. An example of novel compounds with high potency and good therapeutic indices is a group of analogues possessing a 6-arylsulfonyl-2-aminobenzonitrile scaffold (Table 1), for which some rationalization of the structure−activity relationships (SAR) data and the effect of several mutations on the binding affinities has been reported based on the expert, although subjective, interpretation of the crystal structure of a single complex.[1]

Generally speaking, it can be argued that performing a quantitative assessment for a whole series of derivatives from visual examination of a unique complex alone is a daunting task. This is so because the free energy difference that drives ligand−receptor (L−R) association results from a subtle interplay of binding forces that take place within the receptor binding site, usually in the face of competition with water molecules. However, when one such structure is used in conjunction with molecular modeling techniques to compute the interaction energies for a set of ligands that display graded affinity toward a target receptor, quantitative SAR (QSAR) can then be derived.[2] Although some aspects of the modeling procedure itself can also be subjective, the resulting QSAR model can be objectively interpreted and also challenged for predictive ability.

A useful alternative to considering just one global L−R interaction energy for each complex is to decompose this term into a set of van der Waals (vdW) and electrostatic (Ele) contributions emanating from individual receptor amino acids, and project the resulting matrix of energy terms onto a small number of orthogonal "latent variables" (or principal components, PC) using partial least squares (PLS)[3] in a way similar to that used in comparative molecular field analysis (CoMFA).[4] At the end of the procedure, those pairwise interactions between the ligands and individual protein residues that are predictive of activity or binding free energy are selected and weighted according to their importance in the model. Since its inception, this chemometric method, termed comparative binding energy (COMBINE) analysis,[5] has been successfully applied to a variety of biologically relevant targets.[6−9] The details of the whole procedure have been reported elsewhere.[10]

Here, the structure of HIV-1 RT in complex with inhibitor **v** (Table 1), refined at 3.0 Å resolution to an *R* factor of 0.219 (PDB code 1JLQ),[1] was used as a template to model the whole set of complexes by introduction of appropriate modifications into the bound inhibitor.[11] Preferred docking sites for functional groups were evaluated with the program GRID[12] and assisted in the selection of binding modes. AMBER force field[13] parameters (parm99) were assigned to, or consistently derived for,[9,14] inhibitor atoms, and atom-centered RHF 6-31G*//3-21G* charges were calculated using the Gaussian 98 program[15] and the RESP methodology.[16] All of the complexes were energy refined using the SANDER module in AMBER 6.0[17] until the root-mean-square deviation of the energy gradient was less than 0.1 kcal mol^{-1} Å^{-1}. Changes in Ele free energy on molecular association (ΔG_{ele}) were calculated by solving the Poisson−Boltzmann equation[18] and included three separate components: a summation of residue-based solvent-corrected L−R contributions (E_{ele}^{LR}),[6,7,9] and changes in solvation energy of both inhibitor (ΔG_{desolv}^{L}) and receptor (ΔG_{desolv}^{R}) upon complex formation.[19]

Table 1. HIV-1 RT Inhibitory Activity of 2-Amino-6-arylsulfonylbenzonitriles



cmpd	R	IC_{50} (μM)	cmpd	R	IC_{50} (μM)
a	H	6.9	**q**	3-CN	1.8
b	2-OCH_3	1.4	**r**	4-CN	>18
c	3-OCH_3	0.6	**s**	3-CF_3	5.3
d	4-OCH_3	13	**t**	2,5-Cl_2	0.3
e	2-CH_3	4.5	**u**	3,5-Cl_2	0.03
f	3-CH_3	0.2	**v**	3,5- $(CH_3)_2$	0.007
g	4-CH_3	7.3	**w**	3-Br, 5-CH_3	0.003
h	2-Cl	5.9	**x**	3-Cl, 5-CH_3	0.005
i	3-Cl	0.4	**y**	3-OCH_3, 5-CH_3	0.01
k	2-Br	12	**z**	3-OCH_3, 5-CF_3	0.04
l	3-Br	0.2	**dd**	3-$O(CH_2)_3CH_3$,5-CH_3	0.4
m	4-Br	>14			
n	2-F	5.0	**ee**	C_6H_5−R = 1-naphthyl	1.0
p	2-CN	6.0	**ff**	C_6H_5−R = 2-naphthyl	0.03

The optimal dimensionality of each PLS model[20] was determined from the evolution of both the cross-validated correlation coefficient (q^2) and the standard deviation of errors in prediction (SDEP) as a function of the number of PCs extracted (Supporting Information, Table 1). Remarkably, a 3-PC model was able to explain 88.7% of the variance in biological activity for the first 25 inhibitors studied (Figure 1), with very high internal predictive ability (SDEP < 0.4). When this model was used to predict the activity of the most dissimilar compounds, **ee** and **ff**, initially not included in the training set (Table 1), the residuals were only 0.49 and 0.80 log units, respectively (Figure 1). When an expanded set containing all 27 compounds was analyzed, a 4-PC model yielded r^2 and q^2 values of 0.959 and 0.851, respectively, and the most significant PLS coefficients were assigned to just one Ele term representing ligand desolvation and to vdW terms involving about eight amino acids (Figure 2). Of these, P95 and W229 are known to be invariant in HIV-1 RT (and consequently preferential sites for ligand targeting), but the V106A, E138K, and Y181C mutations have differential effects on the potency of compounds **u**−**z** and **ee** as compared to wild-type:[1] roughly 1 order of magnitude in sensitivity is gained in the former case, no significant difference is observed for the E138K mutant enzyme, and a great loss of antiviral activity is apparent against the strain containing the Y181C mutation (more than 3 orders of magnitude). In agreement with these experimental results, the PLS coefficients for the vdW interactions between the inhibitors and the enzyme have opposite signs for Y181 and V106 in the p66 subunit. The chemometric meaning of this finding is that the larger these favorable interactions (negative energy values), the greater the activity (expressed as a pIC_{50}) in the case of Y181 (the largest negative PLS coefficient) but the lesser the activity in the case of V106 (the largest positive PLS coefficient). Therefore,

10.1021/ja038893t CCC: $27.50

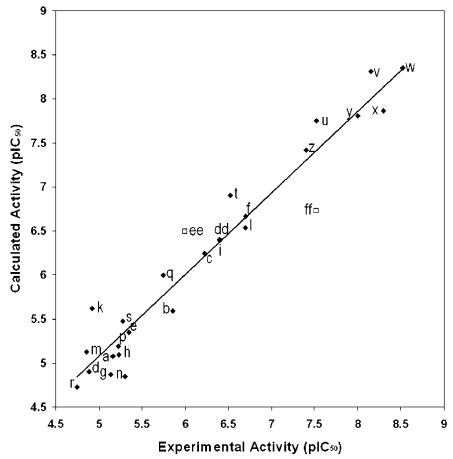


Figure 1. Correlation between calculated and experimental activities for 25 compounds in the training set (◆) and for **ee** and **ff** in the prediction set (□) as obtained in the COMBINE model with three principal components.

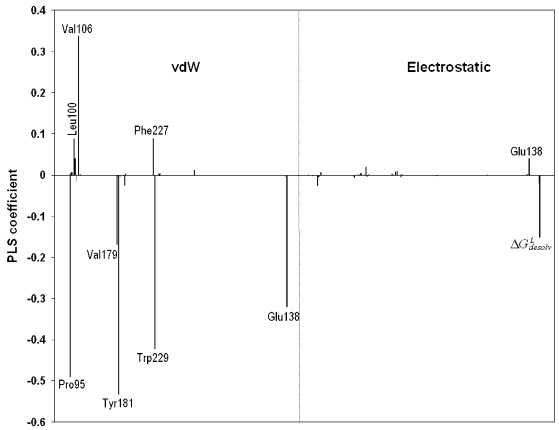


Figure 2. Normalized PLS pseudocoefficients of the van der Waals (left half) and electrostatic (right half) interactions selected by the 4-PC COMBINE model derived for the 27 NNRTIs studied. The most relevant residues have been labeled.

the loss of vdW interactions (Supporting Information, Figure 2) resulting from replacement of V106 with Ala is predicted to be favorable for activity, whereas the corresponding loss upon substitution of Cys for Y181, on the contrary, is predicted to be highly detrimental, in good accord with the experimental evidence. As for E138 in the p51 subunit, the vdW term is likely not to change dramatically when this residue is mutated to a lysine (as opposed to the Ele component).[14]

Regarding the K103N mutation, which is known to be highly deleterious for the activity of NNRTIs, including 2-amino-6-arylsulfonylbenzonitriles, it must be borne in mind that resistance in this mutant has been proposed to arise from stabilization of the closed-pocket form of the enzyme,[21] and also that interaction with K103 is fairly constant throughout the whole series (Supporting Information, Figure 3), due to the limited chemical variation (Table 1). Concerning the L100I mutation, published data for the K103N/L100I double mutant show increases or decreases in activity relative to the already resistant K103N single mutant, depending on the inhibitor. The small positive vdW coefficient assigned to L100 in our model suggests that the effect of the single L100I mutation is not as clear-cut as the others we have discussed.

Extension of this COMBINE methodology to an expanded set containing additional derivatives with substituents exploring the whole of the binding pocket, as well as to other NNRTI series, should provide optimal results and may aid in the design of novel compounds.

Acknowledgment. F.R.-B. is a fellow of the Spanish Ministerio de Ciencia y Tecnología. Financial support from the Spanish CICYT (SAF2000-0153-C02 and SAF2003-07219-C02), the European Commission (QLRT-2000-30291), and the National Foundation for Cancer Research is gratefully acknowledged.

Supporting Information Available: Three figures showing (1) predicted versus experimental activities for the whole set of 27 compounds, (2) selected vdW energies for wild-type and mutated residues, and (3) a profile of the energy variables studied, and a table with chemometric indices (PDF). This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Chan, J. H.; Hong, J. S.; Hunter, R. N., III; Orr, G. F.; Cowan, J. R.; Sherman, D. B.; Sparks, S. M.; Reitter, B. E.; Andrews, C. W., III; Hazen, R. J.; St. Clair, M.; Boone, L. R.; Ferris, R. G.; Creech, K. L.; Roberts, G. B.; Short, S. A.; Weaver, K.; Ott, R. J.; Ren, J.; Hopkins, A.; Stuart, D. I.; Stammers, D. K. *J. Med. Chem.* **2001**, *44*, 1866−1882.

(2) Kubinyi, H., Folkers, G., Martin, Y. C., Eds. *3D QSAR in Drug Design*; Kluwer Academic Publishers: Dordrecht, The Netherlands, 1998.

(3) Wold, S.; Johansson, E.; Cocchi, M. PLS − Partial Least-Squares Projections to Latent Structures. In *3D QSAR in Drug Design. Theory, Methods and Applications*; Kubinyi, H., Ed.; ESCOM Science Publishers B. V.: Leiden, 1993; pp 523−550.

(4) Cramer, R. D., III; Patterson, D. E.; Bunce, J. D. *J. Am. Chem. Soc*. **1988**, *110*, 5959−5967.

(5) Ortiz, A. R.; Pisabarro, M. T.; Gago, F.; Wade, R. *J. Med. Chem.* **1995**, *38*, 2681−2691.

(6) Pérez, C.; Ortiz, A. R.; Pastor, M.; Gago, F. *J. Med. Chem*. **1998**, *41*, 836−852.

(7) Cuevas, C.; Pastor, M.; Pérez, C.; Gago, F. *Comb. Chem. High Throughput Screening* **2001**, *4*, 627−642.

(8) Wang, T.; Wade, R. C. *J. Med. Chem*. **2001**, *44*, 961−971.

(9) Kmunicek, J.; Luengo, S.; Gago, F.; Ortiz, A. R.; Wade, R. C.; Damborsky, J. *Biochemistry* **2001**, *40*, 8905−8917.

(10) (a) Wade, R. C.; Ortiz, A. R.; Gago, F. *Perspect. Drug Discovery Des.***1998**, *9/10/11*, 19−34. (b) Wade, R. C.; Ortiz, A. R.; Gago, F. In *3D QSAR in Drug Design*; Kubinyi, H., Folkers, G., Martin, Y. C., Eds.; Kluwer Academic Publishers: Dordrecht, The Netherlands, 1998; Vol. 2, pp 19−34.

(11) Only those RT residues within 30 Å of the NNRTI binding site were considered, as in ref 14.

(12) Goodford, P. *J. Med. Chem.* **1985**, *28*, 849−857.

(13) 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*, 5179−5197.

(14) Rodríguez-Barrios, F.; Pérez, C.; Lobatón, E.; Velázquez, S.; Chamorro, C.; San-Félix, A.; Pérez-Pérez, M. J.; Camarasa, M. J.; Pelemans, H.; Balzarini, J.; Gago, F. *J. Med. Chem*. **2001**, *44*, 1853−1865.

(15) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A., Jr.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Gonzalez, C.; Challacombe, M.; Gill, P. M. W.; Johnson, B. G.; Chen, W.; Wong, M. W.; Andres, J. L.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. *Gaussian 98*, revision A.11.2; Gaussian, Inc.: Pittsburgh, PA, 2001.

(16) Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. *J. Phys. Chem.* **1993**, *97*, 10269−10280.

(17) http://www.amber.ucsf.edu/amber/amber.html.

(18) (a) Nicholls, A.; Honig, B. *J. Comput. Chem.* **1991**, *12*, 435−445. (b) DelPhi module, Insight II, version 98.0, 1998. Molecular Simulations Inc., 9685 Scranton Road, San Diego, CA 92121-2777.

(19) The interior of the enzyme, the inhibitors, and the complexes was considered a low dielectric medium ($\epsilon = 4$), whereas the surrounding solvent was treated as a high dielectric medium ($\epsilon = 80$) with ionic strength of 0.145 M. Similar results were obtained at zero ionic strength.

(20) Q^2 version 4.5, 2000. Multivariate Infometric Analysis, S.r.l., Perugia, Italy.

(21) (a) Hsiou, Y.; Ding, J.; Das, K.; Clark, A. D., Jr.; Boyer, P. L.; Lewi, P.; Janssen, P. A.; Kleim, J. P.; Rosner, M.; Hughes, S. H.; Arnold, E. *J. Mol. Biol*. **2001**, *309*, 437−445. (b) Lindberg, J.; Sigurðsson, S.; Löwgren, S.; Andersson, H. O.; Sahlberg, C.; Noréen, R.; Fridborg, K.; Zhang, H.; Unge, T. *Eur. J. Biochem*. **2002**, *269*, 1670−1677.

JA038893T