Journal of Computer-Aided Molecular Design **17:** 129–134, 2003.

On the detection of multiple-binding modes of ligands to proteins, from biological, structural, and modeling data

Paul J. Lewis[1,*], Marc de Jonge[1], Frits Daeyaert[1], Luc Koymans[1], Maarten Vinkers[1], Jan Heeres[1], Paul A.J. Janssen[1], Eddy Arnold[2], Kalyan Das[2], Art D. Clark Jr.[2], Stephen H. Hughes[3], Paul L. Boyer[3], Marie-Pierre de Béthune[4], Rudi Pauwels[4], Koen Andries[5], Mike Kukla[6], Donald Ludovici[6], Bart De Corte[6], Robert Kavash[6] & Chih Ho[6]

[1]*Center for Molecular Design, J&JPRD, Janssen Pharmaceutica N.V, Vosselaar, Belgium;* [2]*Center for Biotechnology and Medicine, and Chemistry and Chemical Biology Department, Rutgers University, NJ;* [3]*NIH-NCI HIV Drug Resistance Program, Frederick, MD;* [4]*Tibotec-Virco, Mechelen, Belgium;* [5]*Virology Dept., J&JPRD, Janssen Pharmaceutica N.V, Beerse, Belgium;* [6]*Chemistry Dept., J&JPRD, Spring House, PA, USA*

Received 1 November 2002; Accepted for publication 1 December 2002

Key words: molecular modeling, virology, HIV, X-ray crystallography, drug design, statistical analysis, PLS

Summary

There are several indications that a given compound or a set of related compounds can bind in different modes to a specific binding site of a protein. This is especially evident from X-ray crystallographic structures of ligand-protein complexes. The availability of multiple binding modes of a ligand in a binding site may present an advantage in drug design when simultaneously optimizing several criteria. In the case of the design of anti-HIV compounds we observed that the more active compounds that are also resilient against mutation of the non-nucleoside binding site of HIV1-reverse transcriptase make use of more binding modes than the less active and resilient compounds.

Introduction

Multiple binding modes of ligands to proteins have a dual nature. First, flexible ligands with rotatable fragments may adapt their conformation to a given binding site, at a relatively low cost in energy. Second, both the positions of the backbone and amino acid side chains of proteins can be induced to change their orientation and position in response to ligand binding; both phenomena may appear simultaneously because neither the ligand nor the protein is rigid. The metaphor of a flexible hand fitting into a flexible glove (or, alternatively, a 'handshake') is more applicable here than the rigid lock and key paradigm of Emil Fischer [1].

The dual aspect of multiple binding modes is reflected in many computational protocols for structure-based drug design. For example, some 10 to 20 representative low-energy conformational structures of the ligand may be defined at the start of the computation, each of which deviates from the minimal energy conformation by a relatively small amount (e.g. no more than 5 kcal/mole). Only those structures are retained that can be docked into the rigid binding site without significant steric clashes with the binding site. For each docked conformation the ligand and adjacent amino acid side chains and backbone are relaxed and the overall energy of the complex is minimized. The final result can be a weighted average binding energy over the several conformers that have been docked and optimized. The average may extend over several different binding configurations of the ligand and the protein, depending upon the degree of flexibility that is allowed. A shortcoming of this approach is that it does not permit relatively large displacements of the backbone and the side chains of the binding site, which can be induced by the ligand.

*To whom correspondence should be addressed. E-mail: plewi@prdbe.jnj.com.

Figure 1. X-ray crystal structures of HIV1-reverse transcriptase complexed with the prototype DAPY compound TMC120-R147681 (dapivirine), obtained by E. Arnold and co-workers.

Manifestations of multiple binding

Ligands that have multiple binding modes have been reported [2-6]. Overall, the reports fall into two main categories. In the first one, the same ligand binds in distinct orientations or conformations in the binding site. This has been reported most frequently in separate co-crystallizations of protein and ligand. For example, in one structure the ligand may be in the forward orientation in the binding site, while in another structure the same ligand will bind in the opposite orientation. There are instances, however, where the two binding modes appear together in the same co-crystal [6]. In the second (and more frequent) category, chemically related compounds bind in different orientations. For example, one group of analogs may crystallize in the forward orientation, while the others are oriented backwards. This second category may be a particular instance of the first, in which the different ligand-protein orientations and conformations co-exist in solution, but of which only one crystallizes. It is more likely, however, that apparently small changes in the chemical structure of a ligand, such as the addition of a methyl substituent, can induce considerable changes in the binding, thus offering additional binding modes to the ligand.

Multiple binding modes and drug design

The concept of multiple binding modes may be of importance in drug design. It offers opportunities for optimizing secondary properties such as bioavailability, onset and duration of action, metabolic stabil-

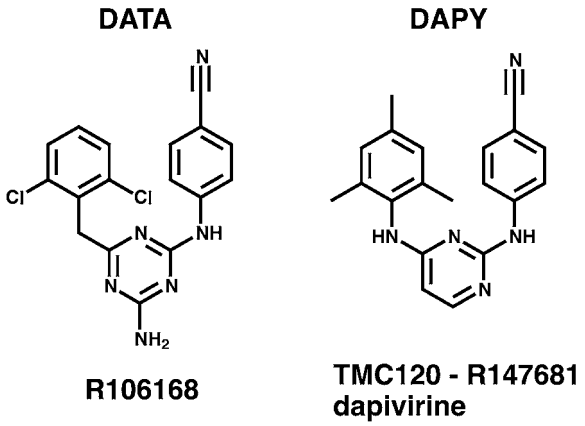

Figure 2. Prototype anti-HIV compounds of the classes of di-aryl triazine analogs (DATA) and di-anilino pyrimidine analogs (DAPY).

ity, etc. A case in point is the design of synthetic neuraminidase inhibitors, which are now available as anti-influenza drugs [7]. Because the binding site is highly polar, the prototype drug zanamivir is also highly polar and is rapidly excreted by the kidney. This requires a special formulation of the drug for powder inhalation. Addition of a lipophilic function to the original compound led to the highly potent and more bio-available drug oseltamavir; the lipophilic portion induced a change at the binding site, which provided for non-polar interactions.

A growing concern in the design of drugs for infectious diseases is the emergence of drug resistance. In the case of tuberculosis, resistance to multiple drugs is developing rapidly, most probably as a result of incomplete and/or inadequate treatment with existing drugs in countries where the resources for proper treatment are lacking. In the case of HIV, resistant strains of the virus may already become detectable after a few weeks of treatment with present-day combination therapy.

It seems therefore important to design drugs that are not only highly potent against the prevailing or wild-type virus, but that also retain their effectiveness against the common resistance mutations. Here we describe a case where multiple binding modes has been important for the design of potent and resilient drugs, i.e. compounds that are highly active against the wild-type and several known drug-resistant strains.

Design of anti-HIV compounds

For the past 15 years the Janssen laboratories, in collaboration with other institutions, have engaged in the search for anti-HIV compounds. TIBO, the first non-nucleoside inhibitor of HIV-1 reverse transcriptase (RT), was discovered in 1987 by screening the Janssen compound library in collaboration with the Rega Institute in Leuven [8]. In 1992, Steitz and co-workers published a structure of RT co-crystallized with nevirapine in the non-nucleoside binding site [9]. Arnold and collaborators reported the structure of RT complexed with TIBO in 1995 [10, 11].

The non-nucleoside binding site of RT is highly flexible and adaptive. In the absence of a ligand, the site forms a compact lipophilic structure, which opens to accommodate a suitable compound [12]. This opening involves a displacement of strands β12-β13-β14 causing repositioning of tryptophan 229 (W229), an essential binding site residue, and the reorientation of the side chains of tyrosines 181 and 188 (Y181 and Y188). This distortion of RT blocks polymerase activity. Important binding determinants for effective NNRTIs include the formation of hydrogen bonds with the peptide backbone atoms of lysine 101 (K101) and lipophilic interactions with the aromatic side chains of Y181, Y188, and W229; a bound NNRTI creates a steric barrier that prevents movement of the indole ring of W229 (and consequently strands β12-β13-β14) back to its location in active HIV1-RT (Figure 1).

In the search for better inhibitors, over 4000 compounds have been synthesized at Janssen during this 15-year period. All of these have been tested for activity against the wild-type virus and the more promising of these have been assayed against a panel of frequently observed mutant strains (including L100I, K103N, Y181C, Y188C). During this period several crystal structures of RT complexed with various non-nucleoside inhibitors have been produced of which some 25 are publicly accessible. (In addition, Arnold and co-workers have produced a series of RT structures co-crystallized with Janssen compounds, which are not yet submitted to the public library.) As a result, we have gathered an important collection of biological, chemical and structural data about this particular protein-ligand interaction. In 1994 Janssen chemists discovered serendipitously the class of di-aryl triazine analogs (DATA) with improved anti-HIV activity when compared to the original TIBO analogs. Subsequently, in 1996 development of the class of di-anilino pyrimidine analogs (DAPY) has been the result of collaboration between medicinal chemistry, crystallography and structural modeling departments (Figure 2) [13–15]. The DAPY class of compounds yielded anti-HIV compounds that are more effective

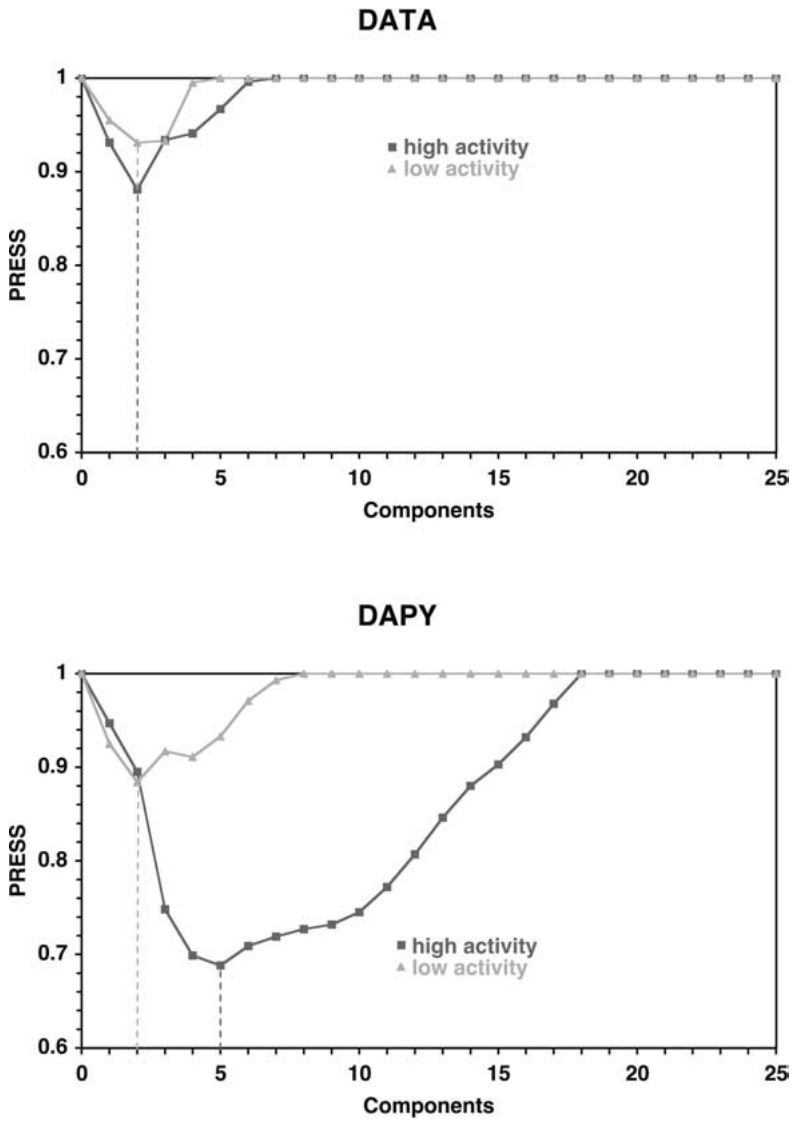

Figure 3. Prediction Error Sum of Squares (PRESS) obtained by means of cross-validated PLS regression as a function of the number of components extracted. PLS regression has been applied to a table of ENERGIES describing anti-HIV compounds in terms of their interaction with the non-nucleoside binding site of HIV1-reverse transcriptase and a corresponding table of ACTIVITIES describing the same compounds in terms of their virological activities against a panel of wild-type and mutant strains. In the case of 140 DATA compounds we find the same 2 binding components for both low and high activity compounds. In the case of 167 DAPY compounds we detect 3 additional binding components in the high activity compounds in excess of the 2 modes in the low activity compounds.

against the wild-type virus and the common drug-resistant mutants than could be obtained within the DATA class. Presently, three members of the DAPY class of compounds are in clinical investigation.

Biological, structural and modeling data

We identified 140 active DATA and 167 active DAPY compounds that could be docked and complex-minimized with 10 to 15 different RT structures. This involved 174 interaction energies with about 90 amino acid residues of the binding site of each complex. The interactions are differentiated according to side chain and backbone of the amino acids, van der Waals or Coulomb potentials and hydrogen binding. We also obtained virological data for each of these 307 compounds in the form of 50 percent effective concentrations (EC_{50}) for inhibiting the wild-type virus and

the four mutant strains mentioned above. (For reasons of analysis, virological activities are expressed as negative logarithms of EC_{50} or pEC_{50}). This yielded a table, called ENERGIES, of 307 anti-HIV compounds described by 174 interactions, together with a corresponding table, called ACTIVITIES, of the same 307 compounds against the five viral strains.

Detection of multiple binding modes

We wanted to ascertain whether multiple binding is important for obtaining high potency of the compounds against both the wild-type and mutant viruses. First, we selected the 140 DATA compounds and divided them into two equal parts, one with average virological activity against the 5 strains at or above the median, and the other with average activity below the median. We applied Partial Least Squares regression (PLS) with cross-validation (using the leave-one-out method) to the selected ENERGIES and ACTIVITIES [16]. In both cases (above and below median activity, respectively) we obtained two significant components, which appeared to be similar with respect to their associated interactions. This suggested, for the DATA compounds, that multiple binding modes are not related to higher or lower activity. When we applied the same analysis to the 167 more potent and resilient DAPY compounds, we obtained 2 significant binding modes for the less active and resilient half of the data set and 5 for the other half at or above the median activity (Figure 3). The 2 modes in the first group are also represented in the 5 modes for the second set. We conclude from this that, at least for the DAPY family, the more active and resilient compounds may take advantage of three additional binding modes, when compared to the less active and less resilient representatives in the DAPY family.

Conclusion

We have analyzed data suggesting that multiple binding of ligands to a protein may be useful in the design of compounds that have to satisfy multiple design criteria, being active against a panel of wild type and mutant targets, for example the non-nucleoside drug binding pocket of wild-type and mutant HIV1-RT.

Multiple binding, when it occurs, may explain why non-nucleoside inhibitors of reverse transcriptase that are highly potent and resilient to mutation are sometimes difficult to crystallize in complex with the target protein. If they crystallize at all, they may also produce only low-resolution X-ray diffraction patterns that do not allow accurate structure determination.

A statistical approach, using Partial Least Squares regression and cross-validation can be used for the detection of the number of significant binding modes from virological, structural and modeling data.

References

1. Fischer, E., Berichte der Deutschen Chemischen Gesellschaft zu Berlin, 27 (1894) 2985.
2. Poulos, T.L. and Howard, A.J., Biochemistry, 26 (1987) 8165.
3. Badger, J., Minor, I., Kremer, M.J., Oliveira, M.A., Smith, T.J., Griffith, J.P., Guerin, D.M.A., Krishnaswamy, S., Luo, M., Rossmann, M.G., McKinlay, M.A., Diana, G.D., Dutko, F.J., Fancher, M., Rueckert, R.R. and Heinz, B.A., Proc. Natl. Acad. Sci. USA, 85 (1988) 3304.
4. Uytterhoeven, K., Sponer, J. and Van Meervelt, L., Eur. J. Biochem., 269 (2002) 2868.
5. Wojtczak, A., Cody, V., Luft, J. and Pangborn, W., Multiple binding modes, http://www.ndsu.nodak.edu/instruct/balaz/bioinf/002.htm
6. Dohnalek, J., Hasek, J., Duskova, J., Petrokova, H., Hradilek, M., Soucek, M., Konvalinka, J., Brynda, J., Sedlacek, J. and Fabry, M., Acta Cryst., D57 (2001) 472.
7. Sham, H.L. and Chen, X., Curr. Pharm. Des., 3 (1997) 159.
8. Pauwels, R., Andries, K., Desmyter, J., Schols, D., Kukla, M.J., Breslin, H.J., Raeymaeckers, A., Van Gelder, J., Woestenborghs, R., Heykants, J., Schellekens, K., Janssen, M.A.C., De Clercq, E. and Janssen, P.A.J., Nature, 343 (1990) 470.
9. Kohlstaedt, L.A., Wang, J., Friedman, J.M., Rice, P.A. and Steitz, T.A., Science, 256 (1992) 1783.
10. Ding, J., Das, K., Moereels, H., Koymans, L., Andries, K., Janssen, P.A., Hughes, S.H. and Arnold, E., Nature Structural Biology, 2 (1995) 407.
11. Das, K., Ding, J., Hsiou, Y., Clark, A.D. Jr., Moereels, H., Koymans, L., Andries, K., Pauwels, R., Janssen, P.A., Boyer, P.L., Clark, P., Smith, R.H. Jr., Kroeger Smith, M.B., Michejda, C.J., Hughes, S.H. and Arnold, E., J. Mol. Biol., 264 (1996) 1085.
12. Jacobo-Molina, A., Ding, J., Nanni, R.G., Clark, A.D. Jr., Lu, X., Tantillo, C., Williams, R.L., Kamer, G., Ferris, A.L., Clark, P., Hizi, A., Hughes, S.H. and Arnold, E., Proc. Natl. Acad. Sci. USA, 90 (1993) 6320.
13. Ludovici, D.W., Kukla, M.J., Grous, P.G., Krishnan, S., Andries, K., de Béthune, M.P., Azijn, H., Pauwels, R., De Clercq, E., Arnold, E. and Janssen, P.A.J., Bioorganic & Medicinal Chemistry Letters, 11 (2001) 2225.
14. Ludovici, D.W., Kavash, R.W., Kukla, M.J., Ho, C.Y., Ye, H., De Corte, B.L., Andries, K., de Béthune, M.P., Azijn, H., Pauwels, R., Moereels, H.E.L., Koymans, L.M.H., de Jonge, M.R., Van Aken, K.J.A., Daeyaert, F.F.D., Lewi, P.J., Das, K., Arnold, E. and Janssen, P.A.J., Bioorganic & Medicinal Chemistry Letters, 11 (2001) 2229.
15. Ludovici, D.W., De Corte, B.L., Kukla, M.J., Ye, H., Ho, C.Y., Lichtenstein, M.A., Kavash, R.W., Andries, K., de Béthune, M.P., Azijn, H., Pauwels, R., Lewi, P.J., Heeres, J., Koy-

mans, L.M.H., de Jonge, M.R., Van Aken, K.J.A., Daeyaert, F.F.D., Das, K., Arnold, E. and Janssen, P.A.J., Bioorganic & Medicinal Chemistry Letters, 11 (2001) 2235.

16. Vandeginste, B.G.M., Massart, D.L., Buydens, L.M.C., de Jong, S., Lewi, P.J. and Smeyers-Verbeke, J., *Handbook of Chemometrics and Qualimetrics*, Elsevier, Amsterdam, 1998, Volume B, p. 368.