Available online at www.sciencedirect.com
SCIENCE DIRECT
Journal of Molecular Graphics and Modelling 24 (2005) 186–194
Journal of Molecular Graphics and Modelling
www.elsevier.com/locate/JMGM

Computational tools for the analysis and visualization of multiple protein–ligand complexes

Sean E. O'Brien [a,b,1,*], David G. Brown [a], James E. Mills [a], Chris Phillips [a], Gregg Morris [a,b]

[a] Department of Medicinal Informatics Structure and Design, Pfizer Global Research and Development, Sandwich, Kent, UK
[b] Accelrys Inc., 10188 Telesis Court, San Diego, CA, USA

Received 12 January 2005; received in revised form 5 July 2005; accepted 10 August 2005
Available online 19 September 2005

Abstract

Modern methods in genomics and high-throughput crystallography have ensured that we have access to a large and rapidly increasing, number of X-ray structures of protein–ligand complexes. A structure-based approach to drug design aims to exploit this information, but current methods are not suited to the examination of the large numbers of complexes available. We present computational tools that analyse and display multiple protein–ligand interactions and their properties in a simplified way.

We illustrate how a novel binding-mode similarity metric is able to cluster 20 ligands complexed to HIV-1 reverse transcriptase into distinct groups. The properties of each cluster are then projected onto a group surface as a series of color gradients. Analysis of these surfaces reveals fundamental similarities and differences in the binding modes of these diverse compounds. In addition, the simplicity of the surface representations facilitates the transfer of information between the crystallographer, computational chemist and the chemist. We also show how two- and three-dimensional (2- and 3-D) similarities can be combined to provide enhanced understanding of 33 factor Xa inhibitor complexes. This methodology has enabled us to identify pharmaceutically relevant relationships between ligands and their binding modes that had previously been hidden in a wealth of data.

Keywords: X-ray structure; Binding mode; Ligand–protein interaction; Surface visualization; HIV-1 reverse transcriptase; Factor Xa

1. Introduction

Modern methods in genome sequencing, proteomics, structural genomics and information technology have provided us with an unprecedented number of new drug targets [1,2]. A rational approach that is ideally placed to exploit these new targets is structure-based drug design. Structure-based methods are an integral part of most industrial drug discovery programs [3] and have produced many successes in recent years [1]. These approaches require determination of the three-dimensional (3-D) structure of the target protein and focus on the nature of its interactions with ligands of interest. An understanding of structure–activity relationships, suggestions for new analogues and generation of novel ideas for ligand moieties are greatly enhanced by knowledge of the structure of the target protein–ligand complex [4]. This information can aid the chemist to optimise compounds by enhancing molecular interactions with the protein to improve potency and selectivity [5].

As of July 1, 2005, over 31,600 structures had been deposited in the Protein Databank (PDB). The vast majority of these are protein and peptide molecules and the bulk of these structures have been determined using X-ray crystallography. Over 5000 of these were deposited in 2004 alone [6]. A significant number of additional structures are also being solved, but not published, due to their proprietary nature. The advent of high-throughput crystallography suggests that the total number of structures produced will swell at a greatly increased rate [2]. Therefore, we are in a situation, where we have vast amounts of structural data at our disposal that can significantly enhance drug discovery.

Having a wealth of information is generally advantageous but it generates the issue of understanding and assimilating all these data. Some academic and industrial research programs now routinely generate large numbers of crystal structures. A researcher working on a project for some time may be able to

* Corresponding author. Tel.: +1 858 875 5103.
E-mail address: sobrien@cylenepharma.com (S.E. O'Brien).
[1] Present address: Cylene Pharmaceuticals, 5820 Nancy Ridge Drive, San Diego, CA, USA.

1093-3263/$ – see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmgm.2005.08.003

absorb new information as it is produced. However, there are likely to be more structures and associated data than can be analysed simultaneously. A computational tool that analyses and displays multiple protein–ligand interactions and their properties, in a simplified way, will greatly enhance learning from the structures. Such software can also be used to exploit the structural information more fully. It may not be enough to analyse and display group properties for all compounds in a dataset. We should also be able to identify the areas of ligand–protein interactions that are similar and different, across a range of diverse chemical entities and also within series.

It is with these goals in mind that we have developed a visualization tool that projects properties from multiple protein–ligand interactions onto a common interaction surface. The surface can be generated from all the atoms in the compounds of interest or any selected subset of atoms. Any user-defined property can then be mapped onto this surface. To determine which molecules are likely to display similar and different interactions with the protein, we have clustered compounds based on their binding-mode similarity. This is a novel measure of similarity and is composed of structural, steric and binding-mode elements. We have developed these methods of X-ray data analysis and illustrate their utility with a set of non-nucleoside reverse transcriptase inhibitors (NNRTIs) and a set of factor Xa (fXa) inhibitors, drawn from the PDB.

NNRTIs are now established as part of multi-drug combinations for treating HIV infection [7,8]. The 'first generation' of drugs, such as nevirapine and delavirdine are susceptible to the effects of a single point mutation within reverse transcriptase (RT) leading to the rapid development of drug resistant virus [9,10]. Progress, however, has been made and the introduction of 'second generation' NNRTIs, such as efavirenz, which show a greater resilience to mutations within reverse transcriptase has led to a new standard of care [11]. The development of drug resistance remains an issue. The RT mutations Leu100Ile, Lys103Asn, Tyr181Cys, Gly190Ser and Met230Leu either singly or in combination result in viruses, which successfully evade all three currently approved agents and efforts to develop compounds with clinical efficacy against these resistant strains are a focus of the pharmaceutical industry [12].

Thrombosis is a major cause of mortality in the western world. Therefore, the control of blood coagulation has become a major target for new therapeutic agents [13]. One attractive approach is the inhibition of fXa, the enzyme directly responsible for thrombin generation [14]. Factor Xa is a well-validated target in thrombosis and clinical data indicate that inhibition of this target gives superior efficacy and safety over thrombin inhibitors [15,16]. In an effort to regulate the blood coagulation cascade, small molecules have been targeted to inhibit this enzyme [17].

2. Methods

2.1. Selection of protein–ligand complex structures

Twenty HIV-1 reverse transcriptase protein–ligand complexes were chosen for analysis. These structures are all of the

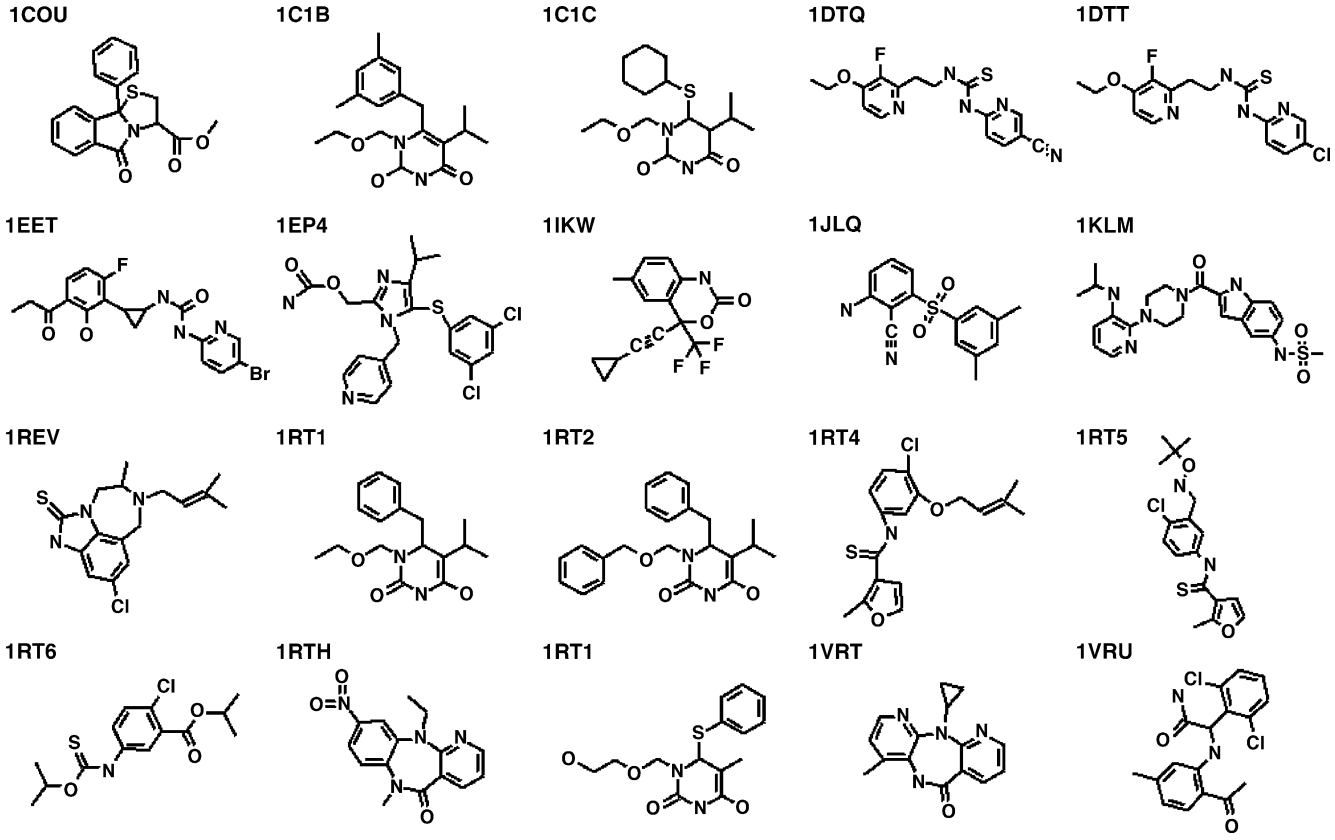

Fig. 1. 2-D structures and PDB codes of 20 HIV-1 reverse transcriptase inhibitors from the PDB.

wild-type viral protein and available from the PDB. This set represents a diverse set of chemical structures as shown in Fig. 1.

The protein structures were overlaid by using a standard least squares algorithm to superpose the α-carbons of the crystal structures onto a reference example. The residues selected to define a reference frame for the RT complexes consisted of eight short segments of structure. These were chosen to define the most structurally invariant regions of the protein active site. The residues are listed as follows: His-A96, Pro-A97, Ala-A98, Gly-A99, Leu-A100; Ser-A105, Val-A106, Thr-A107, Val-A108, Leu-A109, Asp-A110; Val-A179, Ile-A180, Tyr-A181, Gln-A182, Tyr-A183; Asp-A186, Leu-A187, Tyr-A188, Val-A189, Gly-A190, Ser-A191; Phe-A227, Leu-A228, Trp-A229; Tyr-A232, Glu-A233, Leu-A234; Trp-A239, Thr-A240, Val-A241; His-A315, Gly-A316.

Thirty-three human factor Xa inhibitors complexes were chosen from the PDB. The structures of all the compounds are not shown here but six molecules, chosen to illustrate the utility of our similarity measure, are illustrated in Fig. 8. They were overlaid in the same way as the HIV-1 RT ligands, with the residues chosen for superposition listed as follows: Glu-A97, Thr-A98, Tyr-A99; Asp-A189, Ala-A190, Cys-A191, Gln-A192; Val-A213, Ser-A214, Trp-A215, Gly-A216; Cys-A220; Gly-A226, Ile-A227, Tyr-A228.

Atom typing and protonation (at pH 7.0) of the protein–ligand complexes was performed using ADJUSTH (an in-house package—Peter Rose et al.).

2.2. Generation of similarity measure

We calculate a similarity measure between each ligand and every other ligand and then cluster all compounds on the basis of these scores. The hierarchical clustering and subsequent compound display is done in Cerius^{2TM} [18]. We aim to identify compounds that bind in a similar manner. For a given protein with several different ligand-bound structures, overlaid in a common reference frame, a similarity score is constructed that combines three different ways of comparing molecular structures: chemical structure similarity, degree of overlap in 3-D space and a binding site interaction similarity. Each of these measures provides complementary information that enables us to identify the similarities and differences in compound binding modes.

Chemical structural similarity enables us to group chemically similar compounds together. This can be useful in cases where compounds with very similar structure bind in quite different ways [17]. Such ligands will appear in the same region of the cluster dendrogram yet separated by a large (vertical) distance, highlighting crucial differences in the binding modes of chemically related molecules.

The chemical similarity between two compounds is expressed as the Tanimoto similarity score between the Barnard 4096-bit fingerprints of each molecule. The score is calculated for all molecules versus all others under consideration. Thus, a symmetrical $N \times N$ matrix is constructed. This symmetry reflects the nature of the Tanimoto score. It is technically feasible to have an asymmetric similarity coefficient (as exemplified by Bradshaw [19] and based on the ideas of Tversky [20] and Willett et al. [21]) for the Barnard fingerprint overlap and for homogeneity of approach. This will be implemented in the near future. A score of 1 indicates complete 2-D identity, 0 indicates complete dissimilarity.

A steric overlap score indicates whether molecules occupy the same space in the protein pocket, irrespective of chemical make-up. Compounds that exhibit similar binding modes will tend to have a high spatial overlap.

This overlap score comes from the in-house program STERICS, which uses a simple grid-based method to calculate the degree of steric overlap between two molecules. The van der Waals volumes of each ligand are projected onto the same 1 Å grid. The degree of overlap is measured as the proportion of gridpoints in common between the ligands. Again, an $N \times N$ matrix is constructed, but here, element i,j does not equal j,i because STERICS calculates two scores, based on how much molecule B overlaps with molecule A and vice versa. A score of 1 indicates complete overlap, whereas 0 indicates no overlap.

We have constructed a novel binding site interaction similarity score to compare how ligands make contact with the protein's binding pocket. This score is based on a binding site fingerprint that has some correspondence to the Structural Interaction Fingerprint (SIFt) developed by Singh and co-workers [22]. The SIFt scores explicitly identify the interaction type of each common binding site residue with a ligand. As outlined in reference [22], it requires an external analysis of the binding site and the nature of the interaction. Our binding interaction score is more simplistic and only requires the ligand atom types and the protein and ligand atom co-ordinates to be defined.

For all the ligands and associated protein files, we create a list of the shortest contact distances seen in these complexes. The protein points in these contacts are stored as their atom type and residue name. A unique list of the x active site atoms is then generated (x is user-defined—here, we have used 50) that will be used as anchor points to define the fingerprints.

(i) For each compound, the ligand atom nearest to the anchor point is identified. A fingerprint is created depending on its atom type and its distance from the anchor point. At present, the atom types recognized are C, N, O, S and ''other'', although this can be altered. The binning scheme consists of a series of overlapping bins: <2, 1.5–2.5, 2.0–3.0, 2.5–3.5, 3.0–4.0, 3.5–4.5, 4.0–5.0, >5 Å. A molecule having a carbon atom as the nearest atom, 3.1 Å from the anchor point, will have the fingerprint 100000011000. In contrast, a compound with a nitrogen atom as the nearest atom, 2.9 Å from the anchor point, will have the fingerprint 010000110000. There is no explicit bin set for the nearest atom being of type ''other''. The presence of such an atom is recognized by the absence of a C, N, O or S bit being set (i.e. the first four bits being 0000). Hence, the bit length for each point is 12 and not 13 as may be anticipated. The total molecular fingerprint consists of similar binary scores for each anchor point. The fingerprint for a molecule in a binding site containing 50 anchor points will have 600 bits, of which between 100 and 150 bits will be set to 1.

(ii) To generate the similarity matrix, each molecular fingerprint is compared with every other molecular fingerprint and an asymmetric similarity score is generated. The score is defined as the bits set in common by molecules A and B divided by the bits set in molecule A. There is a different score for A compared with B as for B compared with A. Hence, like the steric score, an $N \times N$ matrix is constructed and element i,j does not equal j,i. This approach allows differences in molecular size to be taken into account. A score of 1 indicates complete binding-mode similarity and 0 indicates complete dissimilarity.

Although both the SIFt method and that described above produce fingerprints that attempt to encapsulate binding interaction information, one encodes this by classifying interactions, the other by a comparison of contact distances. The similarity scores that are subsequently generated also differ as we continue to generate an asymmetric matrix of ligand similarity.

In-house testing with large datasets, showing several, diverse binding modes, indicates that, in general, an equal contribution from each of the three measures best highlights the relevant differences in ligand binding modes. Therefore, the three matrices are added together to provide the overall similarity. However, we also have the flexibility to pick and choose which combination of these measures best suits our purposes. The scores are added to an SDF file as descriptors for each compound's binding similarity to all the others.

2.3. Analysis of binding-mode similarity

An SDF file containing all the similarity descriptors is loaded into Cerius^{2TM} and clustered using Ward hierarchical clustering, to generate a cluster diagram. There are an equal number of molecules and similarity descriptors present. Each of the N compounds is described in terms of their binding-mode similarity to each of the N molecules. They are clustered on the basis of these similarity descriptors, and therefore, those molecules, which bind in the most similar fashion will appear close to each other in the hierarchical cluster dendrogram. Cerius^{2TM} enables this cluster diagram to be probed interactively. Clicking on a particular branch or cluster of compounds selects all these molecules, which can quickly be displayed, either overlaid or side-by-side. Other compounds of interest, not present in the particular cluster under examination, can also be displayed. Any data associated with each compound (such as activity or selectivity data) can also be displayed in the Cerius^{2TM} study table to facilitate an understanding of the links between certain structural features and molecular properties.

2.4. Development of visualization tool

Our analysis aims to highlight the similarities and differences between the properties of molecules and groups of molecules. Understanding complex molecular information can be greatly simplified by using a surface representation of the compounds, colored according to atomic or molecular properties. The use of surface properties in ligand and protein–ligand binding analysis is not a novel idea [23–25]. Indeed, many of the capabilities we wish to use already exist in the receptor surface models provided in Cerius^{2TM} [26]. However, we wished to map a number of user-defined properties onto our surfaces and have the additional flexibility of defining any subset of atoms, from any compounds, from which to create the surface.

We have created plug-ins to the ViewerPro [27] interface that enable us to:

(i) import any set of molecules into the ViewerPro along with any set of properties associated with each atom;
(ii) select a subset of these atoms, either by molecule, atom property (an atom is selected if its property value falls within a user designated range) or if it shares the same 3-D volume as any user-defined number of atoms from other molecules;
(iii) create a van der Waals surface around the selected atoms;
(iv) interpolate the property values of the atoms contributing to a surface and represent the property value as a linearly graded color scale (red for the minimum user specified value, through white, then blue for the maximum specified value).

Therefore, we are able to project the properties of any group of atoms onto their combined van der Waals surfaces.

3. Results and discussion

3.1. HIV-1 RT analysis

The 20 HIV-1 RT ligands were clustered as described above (using structure, steric overlap and binding interaction scores) and placed into the six groups shown in Table 1. The final designation of how large a cluster should be is user controlled and the interactive cluster diagram in Cerius^{2TM} enables this decision to be made rapidly. The dendrogram can be seen in Fig. 2. All the compounds in the dataset were displayed within the ViewerPro and surfaces generated around each of the six groups. These surfaces were all colored according to their polarity with red indicating negative, white indicating neutral and blue indicating positive as shown in Fig. 3. The properties of all the atoms contributing to a surface were interpolated onto that surface and the coloring graded to reflect the properties at each point on the surface. All the surfaces in Fig. 3 are shown in the same reference frame.

Table 1
PDB structure codes and associated cluster numbers for HIV-1 RT ligands

Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6
1C0U	1RTH	1RT4	1DTQ	1C1B	1EP4
1IKW	1VRT	1RT5	1DTT	1C1C	1KLM
1JLQ	1VRU	1RT6	1EET	1RT1	
1REV				1RT2	
				1RTI	

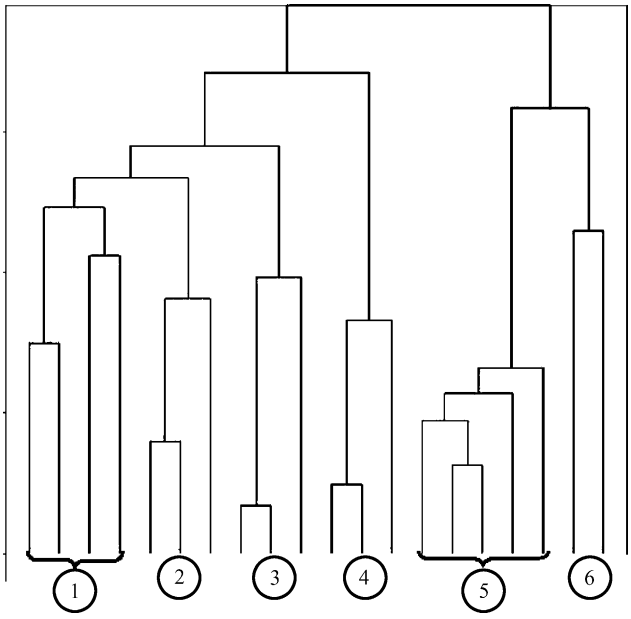

Fig. 2. Dendrogram showing the hierarchical relationships between the clustered HIV-1 RT ligands.

The first thing to notice about each grouping is that they all possess different shapes, although some similarities are obvious. These similarities and differences are considerably more apparent when viewed in three dimensions and a great restriction in producing this paper is the ability to only display two-dimensional (2-D) pictures. Nevertheless, some features can be seen using this representation. All the surfaces in Fig. 3 display an area of positive polarity in the same area except for cluster 2. The green arrow in Fig. 3 indicates the absence of this feature. Upon closer inspection of the structures of these groups, it can be seen that a hydrogen bond donor is present in this position in clusters 1 and 3–6 and not in cluster 2. This can be clearly seen in Fig. 4 in which the surfaces have been removed and a direct comparison between cluster 5 and cluster 2 is shown.

Having identified such differences, their significance can be judged by examining the protein–ligand interactions in this area. Fig. 5 shows that there is a hydrogen bond between the donor and the Lys101 backbone. The example shown is PDB entry 1C1C but this interaction is typical of the compounds not present in group 2. Fig. 6 depicts the absence of a hydrogen bond to the Lys101 backbone in the cluster 2 compounds. However, it does highlight a stabilizing interaction with a water molecule in these structures.

There are many areas of similarity in the surface shape and coloring between groups 3 and 4. However, one area of difference is apparent. In Fig. 7, we show how group 4 compounds fit into the group 3 surface and vice versa. This type of depiction is straightforward in the ViewerPro and very obviously indicates the region shown with the green arrow to be different. In fact, 2 of the group 4 compounds use ring nitrogen atoms to make a hydrogen bond with Glu 138. This interaction is not seen within the rest of the dataset.

Hence, a very quick analysis of these structures has revealed fundamental similarities and differences in the binding modes of these diverse compounds. In addition, the simplicity of the

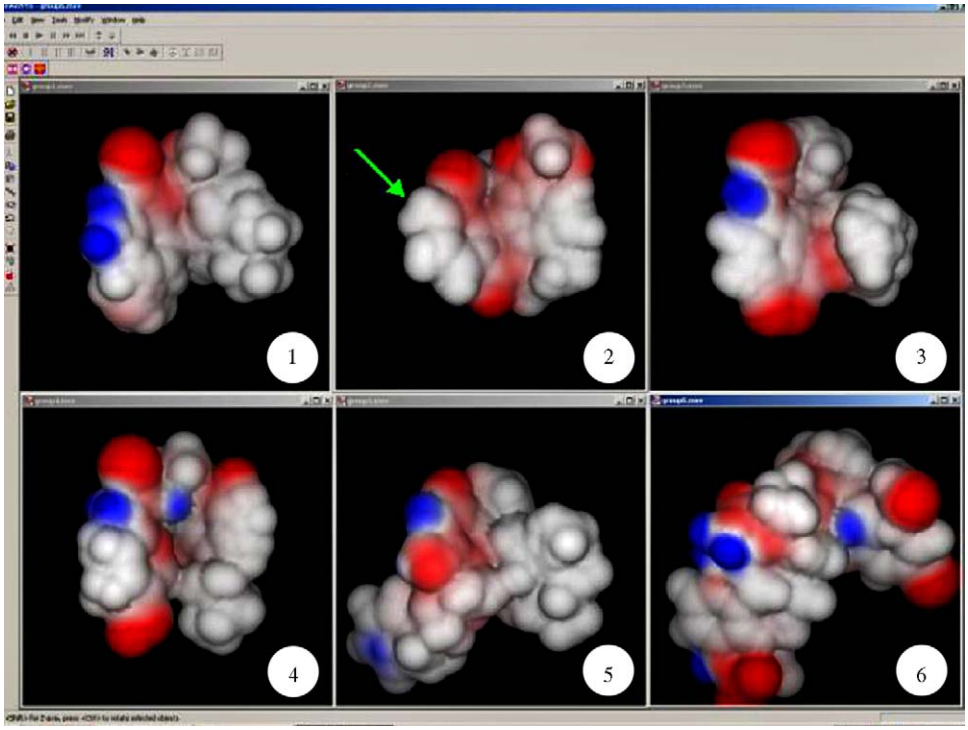

Fig. 3. Surfaces created around the six clusters of HIV-1 RT ligands, colored by polarity. The green arrow indicates a region where the cluster 2 surface differs from all the others.

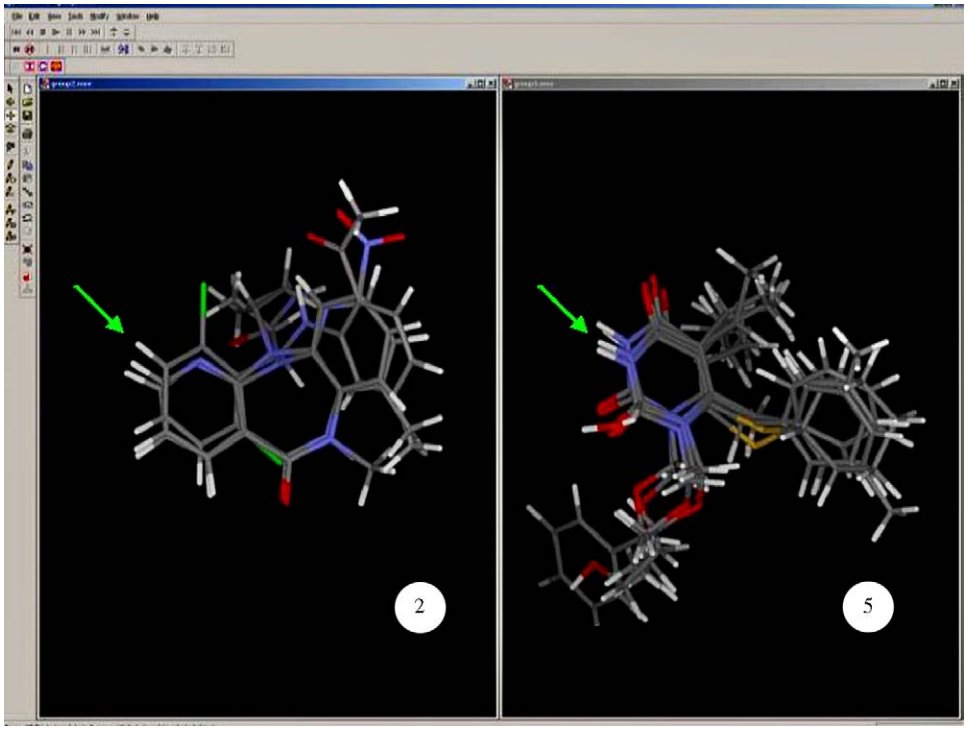

Fig. 4. Ligands in groups 2 and 5, in the same orientation. The green arrow shows where hydrogen-bond donors are present in cluster 5 and absent in cluster 2.

surface representations, with properties mapped on as color gradients, facilitates the transfer of information between the crystallographer, computational chemist and the chemist.

3.2. Factor Xa analysis

The clustering of HIV-1 RT ligands illustrates how compounds with diverse chemical structures can cluster together (groups 1 and 2). However, this example does not illustrate how molecules with very similar structures can be clustered in different areas of the dendrogram due to their binding modes. The clustering methodology, as applied to fXa, however, does illustrate this behaviour. Of the six fXa inhibitors shown in Fig. 8, the first four (1F0R, 1F0S, 1NFX and 1NFW) have closely related structures and would be expected to share the same binding mode [17]. The identities of all the fXa inhibitor complexes used for this analysis are contained in Table 2.

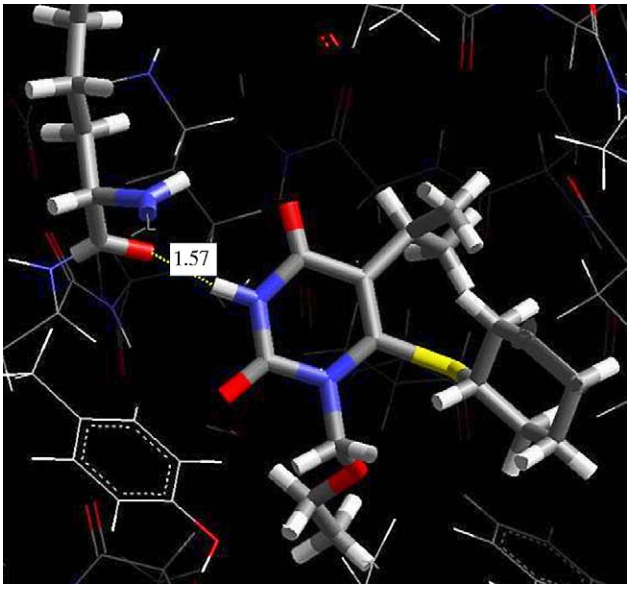

Fig. 5. The hydrogen bond to the Lys101 backbone by 1C1C (non-group 2 compound). The oxygen–hydrogen distance is shown.

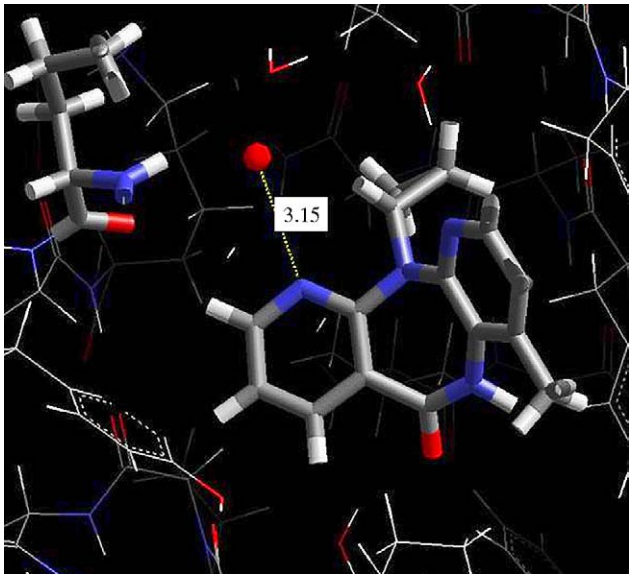

Fig. 6. The absence of a hydrogen bond to the Lys101 backbone by 1VRT (group 2 compound). The stabilizing interaction with a neighboring water molecule is highlighted. The oxygen–nitrogen distance is shown.

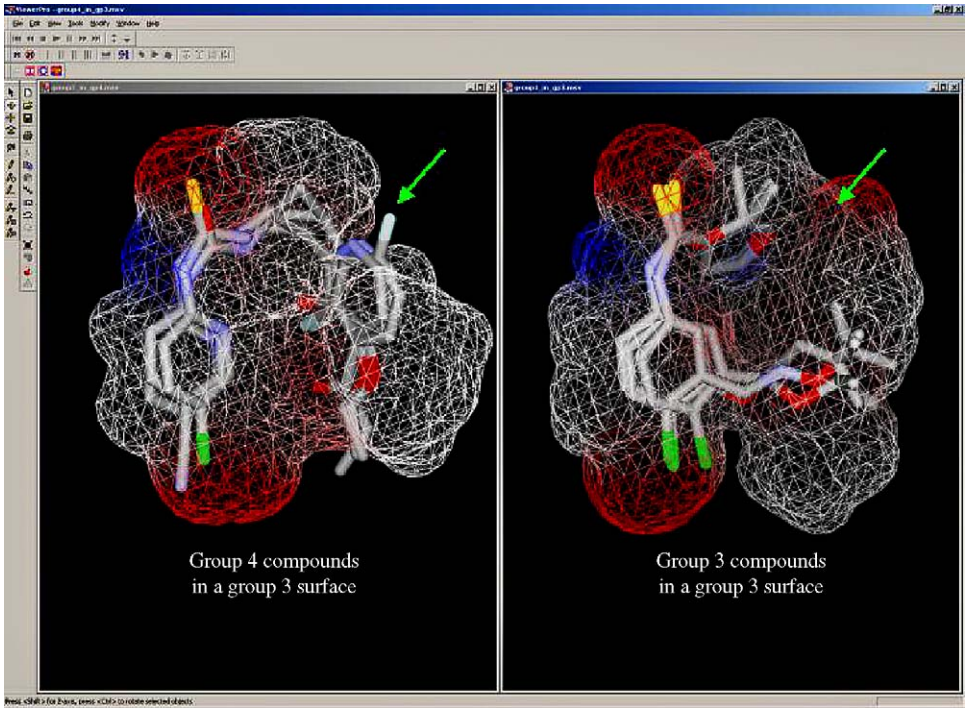

Fig. 7. Displaying compounds in the combined surface generated from a similar group highlights the differences indicated by the green arrow. This region is responsible for a hydrogen bond to Glu138 in group 4. This is absent in group 3 compounds.

The similarity measure used to calculate the dendrogram in Fig. 9 contains the three elements described above—structural similarity, steric overlap and binding interaction score. It can be seen that compounds 1F0R and 1F0S are most closely related to 1NFW and then to 1NFY. However, they are not tightly clustered in the dendrogram. Upon closer inspection, this highlights that these molecules are chemically similar, but they bind in a different manner.

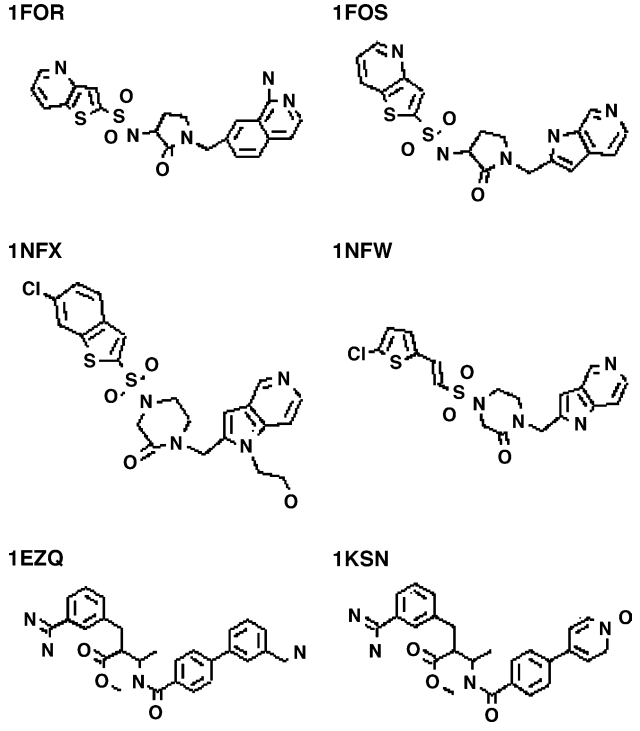

Fig. 8. 2-D structures and PDB codes of six factor Xa inhibitors from the PDB.

Having identified some chemically similar compounds that bind in different modes, we can tailor our similarity measure to probe the dataset further. The dendrogram in Fig. 10 was constructed using a modified binding-mode similarity metric. This one did not include the structural similarity element. A similarity metric that only includes steric overlap and binding interaction score was used. In this diagram, it is clear that compounds 1F0R and 1F0S do not bind like 1NFX and 1NFW. In fact, their binding mode is much more similar to that of 1KSN and 1EZQ. These latter two molecules have a distinctly different chemical structure to the other four compounds (Fig. 8). The actual binding orientations of all six of these structures are shown in Fig. 11. It is clear that 1NFX and 1NFW bind in a 'reverse' mode [17] to that of 1F0R and 1F0S. This illustrates how a 3-D binding-mode similarity metric provides information that is complementary to 2-D structural measures.

Table 2
PDB structure codes for 33 factor Xa inhibitors

1EZQ	1G2M	1IQI	1KSN	1MQ5	1QL7
1F0R	1IOE	1IQJ	1KYE	1MQ6	1QL9
1F0S	1IQE	1IQK	1LPG	1NFU	1XKA
1FAX	1IQF	1IQL	1LPK	1NFW	
1FJS	1IQG	1IQM	1LPZ	1NFX	
1G2L	1IQH	1IQN	1LQD	1NFY	

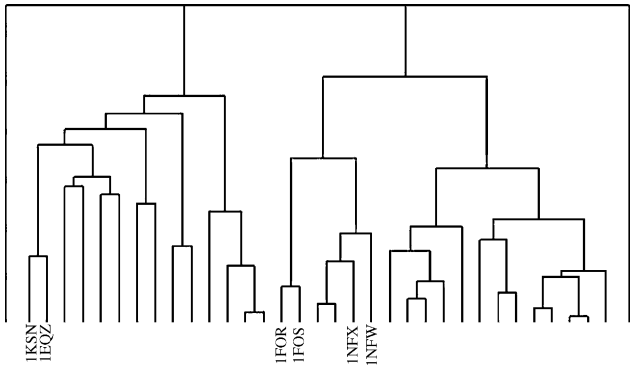

Fig. 9. Dendrogram showing the hierarchical relationships between the factor Xa inhibitors. The binding-mode similarity metric includes a chemical structure similarity measure as well as steric overlap and binding interaction scores. The positions of compounds in Fig. 8 are shown.

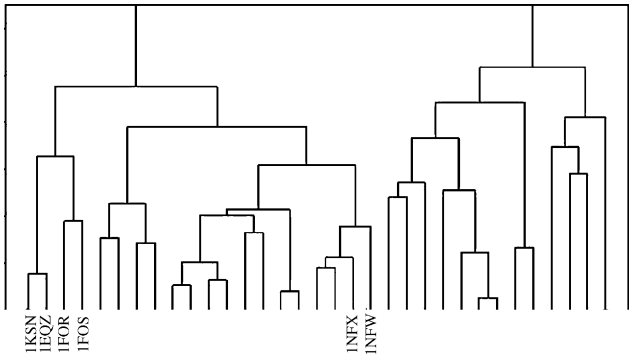

Fig. 10. Dendrogram showing the hierarchical relationships between the factor Xa inhibitors. The binding-mode similarity metric only includes steric overlap and binding interaction scores. The positions of compounds in Fig. 8 are shown.

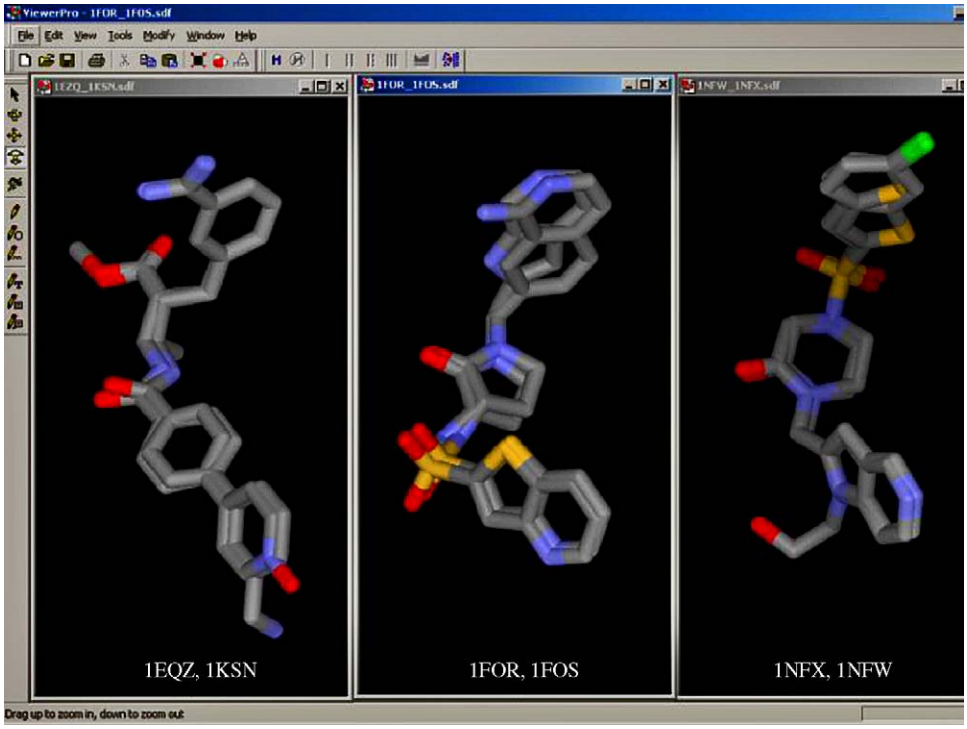

Fig. 11. 2-D structures and PDB codes of six factor Xa inhibitors from the PDB. Compounds 1F0R and 1F0S cluster with 1EZQ and 1KSN when steric overlap and binding interactions are considered.

The inclusion (or removal) of different similarity elements enables us to discover several relationships between molecules and their binding modes.

In this paper, we have illustrated our approach with only a small set of molecules and simple atomic properties for clarity. However, we have successfully used these techniques for more complex analysis of significantly larger in-house datasets. This methodology has enabled us to identify pharmaceutically relevant relationships between ligands and their binding-mode similarities that had previously been hidden in the wealth of three-dimensional data provided by large numbers of protein–ligand structures.

4. Conclusion

The combination of a discriminating similarity score and the interactive nature of the cluster diagram create a powerful tool for structural data analysis. When allied to the display of group properties on combined surfaces, the exploration of X-ray structural data is greatly enhanced.

These techniques can be used as simple learning tools that quickly highlight the structural similarities and differences within a dataset or as methods for more sophisticated analyses of multiple X-ray structures. These are not computational methods to replace the detailed examination of protein–ligand

interactions but they can suggest relationships between chemical structures and their binding modes that may not have been otherwise apparent. This is an approach that can assist in the digestion of the large quantities of X-ray structural data that we have at present and that will increase in the future.

Acknowledgements

We would like to express our thanks for the significant help in modifying the ViewerPro to John Clark and Gerard Lanois of Acclerys Inc.

References

[1] A.C. Anderson, The process of structure-based drug design, Chem. Biol. 10 (2003) 787–797.
[2] T.L. Blundell, H. Jhoti, C. Abell, High-throughput crystallography for lead discovery in drug design, Nat. Rev. Drug Discov. 1 (2002) 45–54.
[3] V. Mountain, Astex, structural genomix, and syrrx, Chem. Biol. 10 (2003) 95–98.
[4] J. Greer, J.W. Erickson, J.J. Baldwin, M.D. Varney, Application of three-dimensional structures of protein target molecules in structure-based drug design, J. Med. Chem. 37 (1994) 1035–1054.
[5] T.L. Blundell, Structure-based drug design, Nature 384 (1996) S23–S26.
[6] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne, The protein data bank, Nucleic Acids Res. 28 (2000) 235–242.
[7] E. De Clerq, The role of non-nucleoside reverse transcriptase inhibitors (NNRTIs) in the therapy of HIV-1 infection, Antiviral Res. 38 (1998) 153–179.
[8] B.G. Gazzard, Efavirenz in the management of HIV infection, Int. J. Clin. Pract. 53 (1999) 60–64.
[9] D. Richmanan, C.K. Shih, I. Lowry, J. Rose, P. Prodanovich, et al. Human immunodeficiency virus type 1 mutants resistant to nonnucleoside inhibitors of reverse transcriptase arise in tissue culture, Proc. Natl. Acad. Sci. 88 (1991) 11241–11245.
[10] R.F. Schinazi, B.A. Larder, J.W. Mellors, Mutations in retroviral genes associated with drug resistance, Int. Viral News 4 (1996) 95–107.
[11] K.A. Sepkowitz, AIDS—the first 20 years, New Engl. J. Med. 344 (2001) 1764–1770.
[12] A.M. Doherty (Ed.), Annual Reports in Medicinal Chemistry, Elsevier Press, San Diego, 2003.
[13] D. Leung, G. Abbernante, D.P. Fairlie, Protease inhibitors: current status and future prospects, J. Med. Chem. 43 (2000) 305–341.
[14] W.R. Ewing, H.W. Pauls, A.P. Spada, Progress in the design of inhibitors of coagulation factor Xa, Drugs Future 24 (1999) 771–787.
[15] N.A. Prager, D.R. Abendschein, C.R. McKenzie, P.R. Eisenberg, Role of thrombin compared with factor Xa in the procoagulant activity of whole blood clots, Circulation 92 (1995) 962–967.
[16] B. Kaiser, Thrombin and factor Xa inhibitors, Drugs Future 23 (1998) 423–436.
[17] S. Maignan, J.-P. Guilloteau, Y.M. Choi-Sledeski, M.R. Becker, W.R. Ewing, et al. Molecular structures of human factor Xa complexed with ketopiperazine inhibitors: preference for a neutral group in the S1 pocket, J. Med. Chem. 46 (2003) 685–690.
[18] Accelrys Inc., Cerius2, 4.91, San Diego, CA, 2004.
[19] J. Bradshaw, Introduction to the Tversky similarity measure, Daylight UGM Meeting, Laguna Beach, CA, 1997.
[20] A. Tversky, Features of similarity, Psych. Rev. 84 (1977) 327–352.
[21] P. Willett, J.M. Barnard, G.M. Downs, Chemical similarity searching, J. Chem. Inf. Comp. Sci. 38 (1998) 983–996.
[22] Z. Deng, C. Chuaqui, J. Singh, Structural interaction fingerprint (SIFt): a novel method for analyzing three-dimensional protein–ligand binding interactions, J. Med. Chem. 47 (2004) 337–344.
[23] N. Stiefl, K. Baumann, Mapping property distributions of molecular surfaces: algorithm and evaluation of a novel 3D quantitative structure–activity technique, J. Med. Chem. 46 (2003) 1390–1407.
[24] K. Hasegawa, S. Matsuoka, M. Arakawa, K. Funatsu, New molecular surface-based 3D-QSAR method using Kohonen neural network and 3-way PLS, Comp. Chem. 26 (2002) 583–589.
[25] J. Polanski, B. Walczak, The comparative molecular surface analysis (COMSA): a novel tool for molecular design, Comp. Chem. 24 (2000) 615–625.
[26] M. Hahn, Receptor surface models. Part 1: definition and construction, J. Med. Chem. 38 (1995) 2080–2090.
[27] Accelrys Inc., ViewerPro 5.7, San Diego, CA, 2004.