

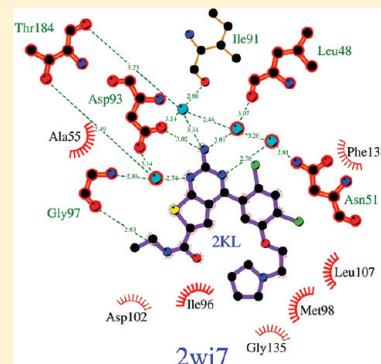
# LigPlot+: Multiple Ligand–Protein Interaction Diagrams for Drug Discovery

Roman A. Laskowski<sup>\*†</sup> and Mark B. Swindells<sup>‡</sup>

<sup>†</sup>European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SD, United Kingdom

<sup>‡</sup>Ebisu, 12 High Street Easton on the Hill near Stamford, PE9 3LR, United Kingdom

**ABSTRACT:** We describe a graphical system for automatically generating multiple 2D diagrams of ligand–protein interactions from 3D coordinates. The diagrams portray the hydrogen-bond interaction patterns and hydrophobic contacts between the ligand(s) and the main-chain or side-chain elements of the protein. The system is able to plot, in the same orientation, related sets of ligand–protein interactions. This facilitates popular research tasks, such as analyzing a series of small molecules binding to the same protein target, a single ligand binding to homologous proteins, or the completely general case where both protein and ligand change.



## INTRODUCTION

Despite well-known limitations, the structure determination of a ligand in complex with its target protein is generally regarded as a highly valuable piece of information for helping to understand ligand–target complementarity and possibly direct subsequent design strategy. Detailed inspection of available 3D structures will provide the fullest interpretation, but these are often difficult to investigate quickly, even for experts; while potential users from other disciplines, such as medicinal chemistry, need time to first acquaint themselves with 3D coordinate visualization software. As a consequence, there is a barrier to sharing valuable 3D structure information, whether determined by crystallography or NMR or predicted from protein homology modeling and small molecule docking techniques.

The original LIGPLOT program<sup>1</sup> was written to focus on specific interactions, most commonly between ligand and protein, including interactions mediated via water molecules or via a specific residue, such as the Asp interaction in the Ser-His-Asp catalytic triad of serine proteases. Other types of interactions could also be plotted, such as dimerization surfaces and specific domain–domain interactions. The idea behind the original program was to take the 3D coordinates of a protein–ligand complex, calculate the ligand–protein hydrogen bonds and non-bonded contacts, and then “flatten out” the ligand and interacting protein residues to give a clear 2D diagram with minimal atom clashes and overlapping of bond lines. The hydrogen-bond and hydrophobic interactions were automatically calculated by the HBPLUS program<sup>2</sup> with the flattening and optimization steps then being performed on that data.

Over the years LIGPLOT has been extensively used by researchers publishing new structures. LIGPLOT diagrams have

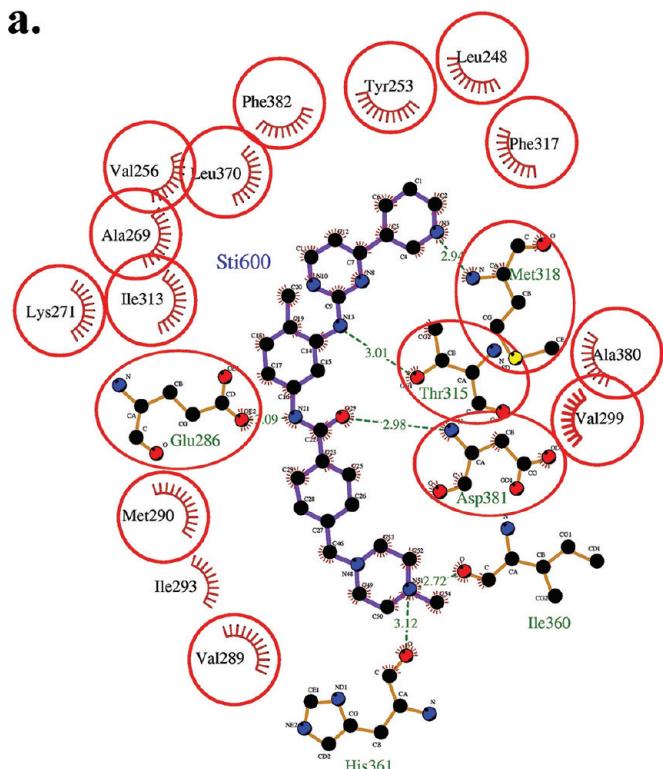
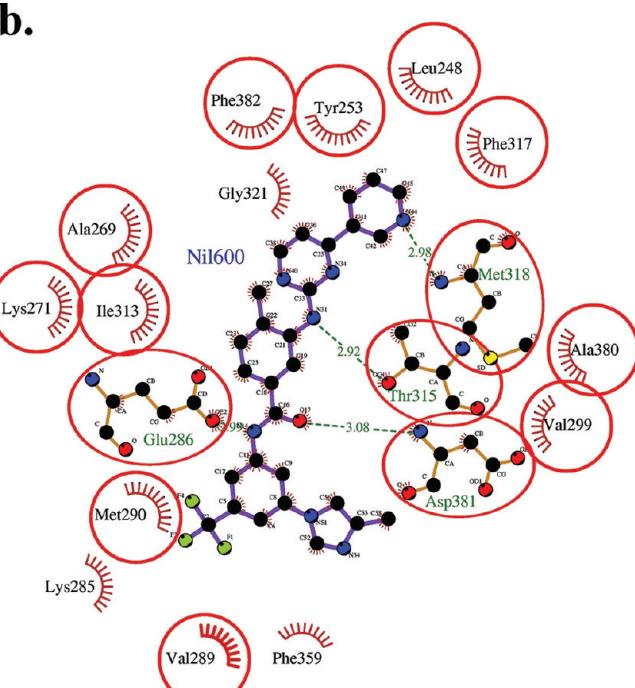
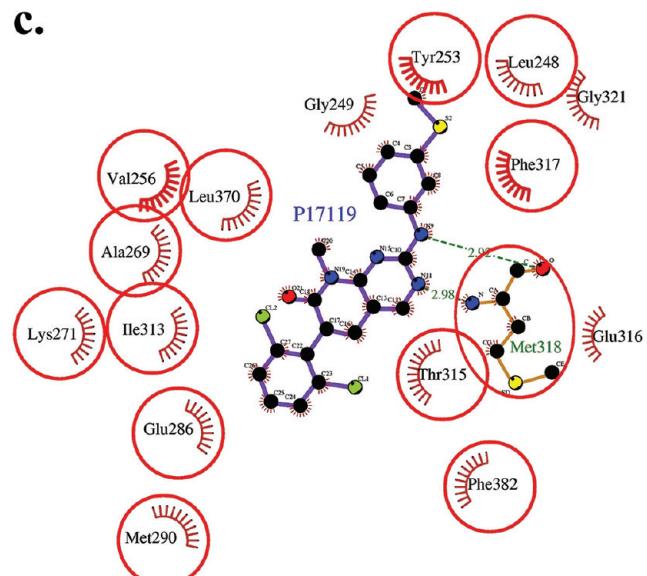
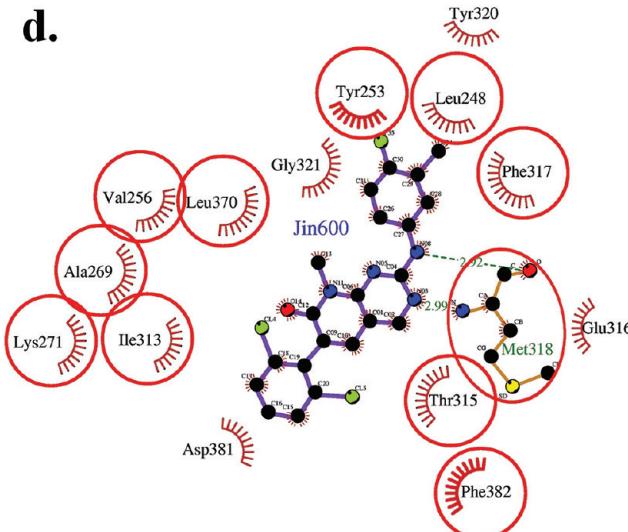
also been incorporated into the PDBsum database<sup>3</sup> to illustrate ligand–protein interactions for all structural models in the Protein Data Bank (PDB).<sup>4</sup> A java helper program, called LigEd, was subsequently added to allow interactive editing of LIGPLOT diagrams. This made it possible to move the components of the diagram about on screen, to rotate them about selected atoms, flip them about selected bonds, change their colors, and edit their text labels.

Since the original publication of LIGPLOT, other algorithms have been proposed based on similar ideas. One system by Steirand et al.<sup>5</sup> first generates the 2D representations of ligand and protein side chains using the Structure Diagram Generation Library (LibSDG)<sup>6</sup> and then lays these out on the page in a manner that minimizes clashes and overlaps. Hydrogen bonds are calculated by FlexX<sup>7</sup> and, as in LIGPLOT, are shown as dotted lines. Hydrophobic contacts are represented as spline curves that outline the hydrophobic parts of the ligand and are labeled by the contacting amino acid residues from the protein.

Both the original LIGPLOT, and the approach proposed by Steirand et al.,<sup>5</sup> concentrated on plotting a single ligand complex. As more ligand complexes have become available comparative analysis has become important. The Molecular Operating Environment (MOE) program<sup>8</sup> has tried to address this within their system by allowing more than one complex to be statically plotted in a comparable orientation. In their implementation, each ligand is flattened in 2D to preserve as close a representation of its 3D structure as is possible in 2D. However, for comparing interactions from related structures, the system relies on an initial

Received: May 23, 2011

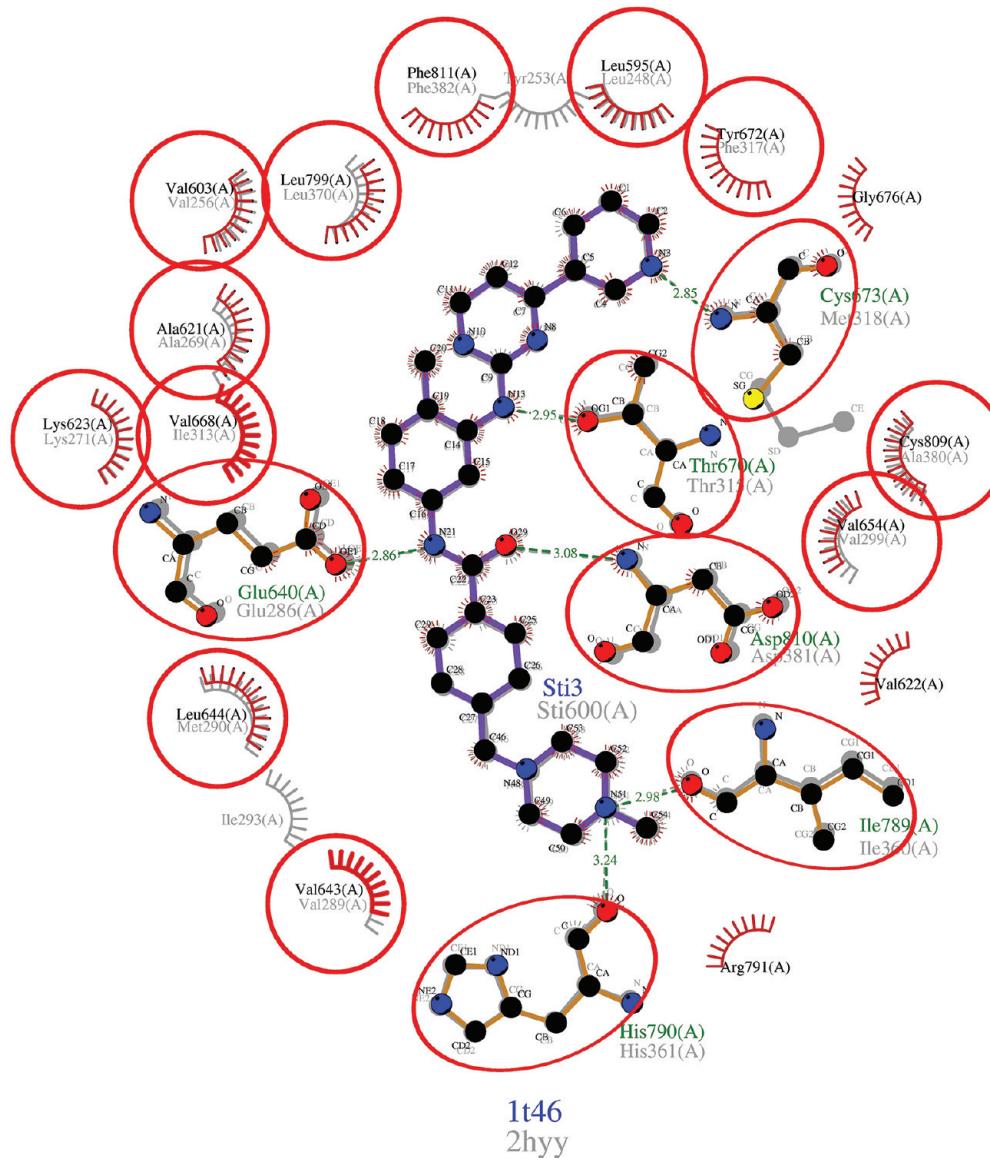
Published: September 15, 2011

**a.****b.****c.****d.**

**Figure 1.** Ligand–protein interaction diagrams for four binding sites of the same protein (human c-Abl tyrosine kinase) each with a different ligand molecule bound. The ligands are: (a) imatinib (PDB entry 2hy), (b) nilotinib (3cs9), (c) PD173955 (1m52) and (d) PD180970 (2hzi). The plots were generated by LigPlot+ in the order given, with each subsequent plot being automatically fitted to the first (2hy). The red circles and ellipses in each plot indicate protein residues that are in equivalent 3D positions to the residues in the first plot. Hydrogen bonds are shown as green dotted lines, while the spoked arcs represent residues making nonbonded contacts with the ligand.

global superposition of the 3D structures, so that the equivalences between ligands and interacting residues can be identified. This step requires familiarity with their 3D visualization programs.

Furthermore the positioning of ligands and labels are determined by the program and cannot be adjusted by the user. A similar approach has recently been proposed by Schrödinger



**Figure 2.** A superposed plot of imatinib bound to human c-Kit tyrosine kinase (PDB entry 1t46) and to human c-Abl tyrosine kinase (2hy). The former plot is shown in full color and is overlaid on the 2hy plot, which can be seen here and there in gray. The red circles and ellipses identify the equivalent residues in the two 3D structures. The plots were overlaid automatically by LigPlot+, but the residue labels have been manually repositioned for clarity.

LLC, using a set of scripts to generate a static ligand interaction diagram.<sup>9</sup>

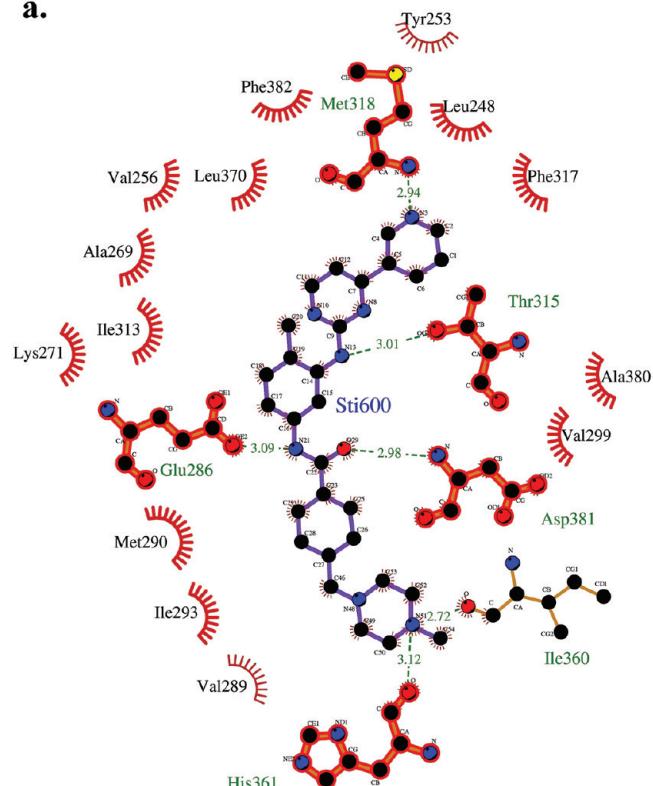
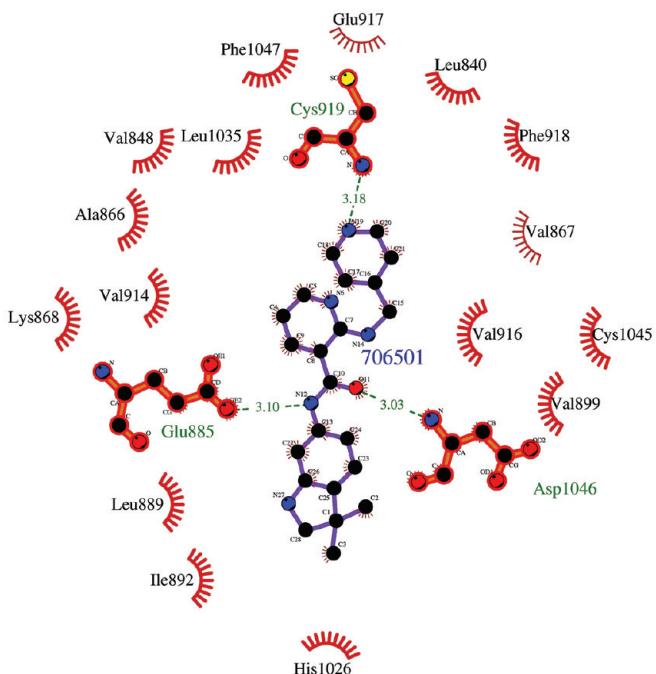
Here we describe LigPlot+, a complete reworking of the original LIGPLOT program with a new graphical user interface (GUI). The program allows the 2D representation of multiple ligand–protein complexes in a simple and automated manner, without the need of prealigning the protein structure components or using 3D viewers. The diagrams can be manually edited at any stage in the process. Thus, for example, edits made to the first diagram form the basis for the layout of subsequent diagrams.

The ability to display multiple ligand complexes makes LigPlot+ readily applicable to a wide range of research considerations, including a single ligand binding to different proteins and multiple ligands binding to the same protein as well as more general pharmacogenomic concepts where both ligand and protein differ. It can be used for experimentally determined structural models,

homology models, or docking results, where it can be used to analyze ensembles of ligand–docked complexes. The fitting and alignment of the 3D structures is entirely automated, so all the user needs to do is select the relevant structures in PDB format. For each subsequent plot, the program overlays them, one after another, automatically highlighting which residue positions are in equivalent positions in the superposed 3D structures.

## METHODS

**LigPlot+ Interface and Integration with Third Party Visualization Programs.** The LigPlot+ graphical user interface is written in Java. Once the first, or any subsequent plot, appears on screen, it can be edited in a number of ways. The ligand molecule and protein side chains can be moved, rotated about any

**a.****2hyy****b.****3efl**

**Figure 3.** A plot showing the binding of different molecules by two homologous proteins. The left-hand plot shows imatinib bound to human c-Abl tyrosine kinase (PDB entry 2hyy), while the right-hand plot shows the binding of motesanib to the kinase domain of vascular endothelial growth factor receptor 2, VEGFR2, (3efl). Equivalent hydrogen-bonding residues are shown with a red border, while equivalent residues involved in nonbonded contacts are shown by the thick, red spoked arcs. Although the overall sequence identity of the two proteins is around 35%, the similarity of their binding sites is much higher, which allows LigPlot+ to identify equivalent residues (and consequently equivalent ligand atom positions) and allow it to produce two plots showing the ligand-binding similarities of these two structures.

atom, or flipped about any bond. Colors can be altered, text items can be edited, and various depictions can be switched on or off.

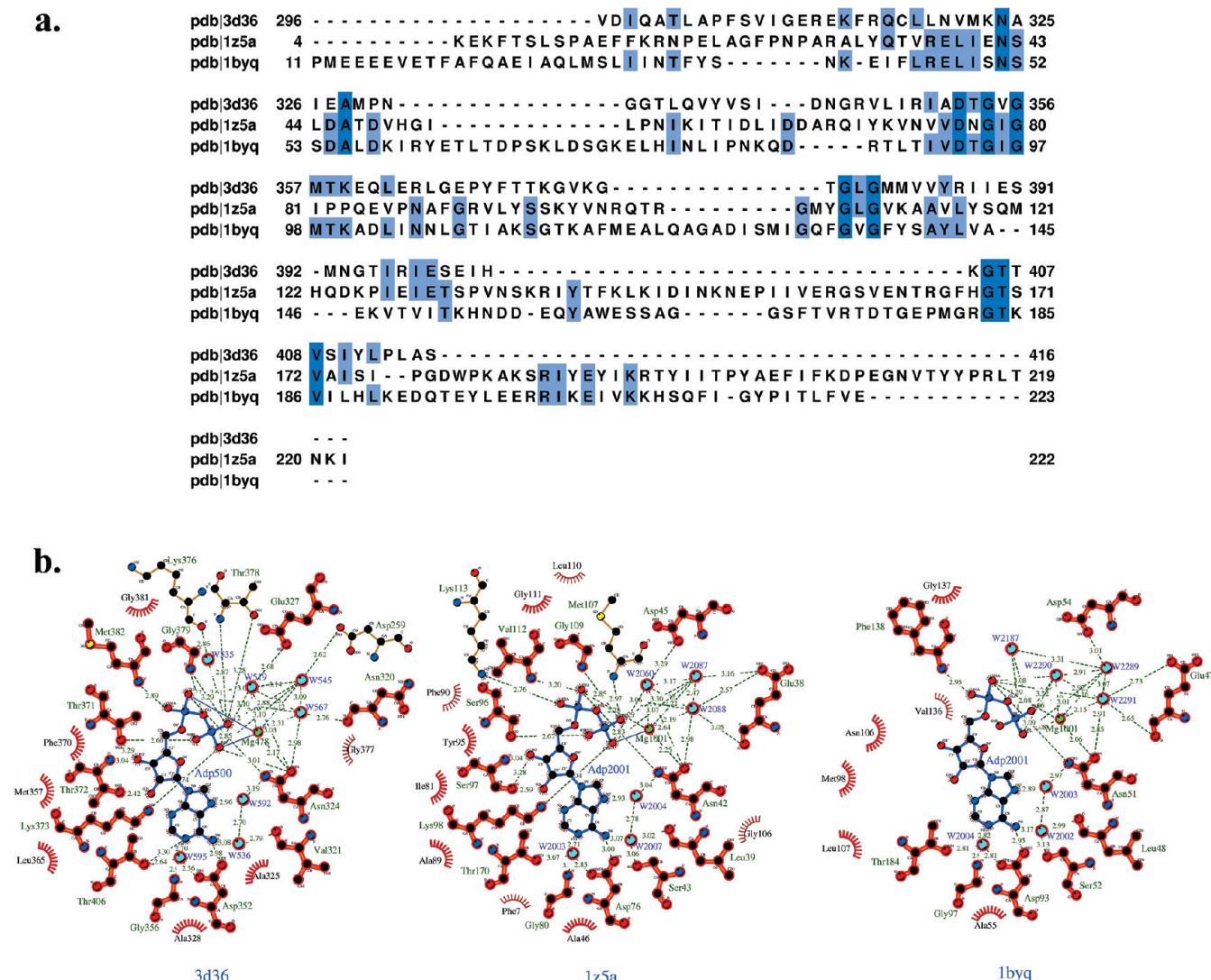
Plots are overlaid based on structural similarities identified by the algorithm described later. Residues in equivalent positions in the 3D structures after superposition are placed in equivalent positions and orientations on the 2D plot. The ligands are positioned so as to optimize the number of structurally equivalent atoms superposed. Equivalent residue positions, which need not be of the same amino acid type, are superposed and highlighted. When the plots are superposed, the currently 'active' plot is shown in color on top and is the plot that can be edited. Inactive plots are greyed out underneath but can be brought to the front by either clicking on the plot's label or by clicking on any greyed out element. A toggle button allows the screen to be split, with each plot shown separately, or for the plots to be merged back together again. The plots can be written to a PostScript file for printing either as shown on screen or each on a separate page. In principle, any number of plots can be superposed, although the complexity rapidly escalates. The PostScript plots can then be converted to any image format using some appropriate image processing software.

If either RasMol<sup>10</sup> or PyMOL<sup>11</sup> are installed on the user's system, the 3D coordinates of the corresponding interactions can be viewed by clicking the appropriate button. This is achieved

through a script that restricts the display to information seen on the corresponding LigPlot+ diagram.

**Data Input From PDB Files.** The primary input to the program is a PDB-format file. For released PDB entries, this is most conveniently accessed via the four-character PDB identifier. The file is then retrieved either from a predefined location on the user's system or via the web from one of the wwPDB ftp sites.<sup>12</sup> Alternatively, any proprietary PDB-format file on the user's system can be uploaded via a browse option. The program first scans the file to identify the ligands and shows the user a list of any it finds. The user needs only to select the ligand of interest for the program to set off and generate the plot. If the ligand is not identified by the program, it can be specified by entering an appropriate residue range. Multiple interacting ligands in a common pocket can also be selected using the range option in order to plot together. It is important that proprietary structures follow the published PDB format,<sup>13</sup> particularly when multiple ligands are present and when the researcher wishes to show more than one ligand on the same diagram.

**Finding Structural Equivalence.** A triage of methods is used to overlay individual ligand–protein complexes. First, the program assumes the two proteins are related and aligns their sequences using a simple Needleman and Wunsch<sup>14</sup> alignment. The alignment forms the basis of a 3D superposition of the two



**Figure 4.** Plots fitted on the basis of a structural superposition of three distantly related protein structures. The three proteins are: histidine kinase B (PDB code 3d36), DNA gyrase B (1z5a), and HSP90 (1byq). All bind ADP. (a) The sequence alignment obtained from the CATHEDRAL multiple-structure alignment. Identical residues in the three proteins are highlighted in dark blue, while residues that are identical in two of the three structures are colored light blue. (b) The LigPlot+ diagrams of the ADP binding sites from the three proteins generated on the basis of the structural alignment. Equivalent residues and water molecules are highlighted with a red outline. Only conserved waters have been included and are colored cyan.

structures. If the sequence identity is too low (i.e., <30% or alignment overlap < 30 residues) the program tries to perform a sequence alignment on just the binding site residues, as these are often more highly conserved than the sequence as a whole. These residues are identified from the HBPLUS output, and their names and numbers are listed in sequence. Any gaps in either sequence are filled by a single dummy residue. For example, if the interacting residues in the first structure are Leu248, Tyr253, and Val256, the binding site's sequence is taken to be LXYXV, where each 'X' residue is the dummy residue corresponding to one or more missing residues in the sequence.

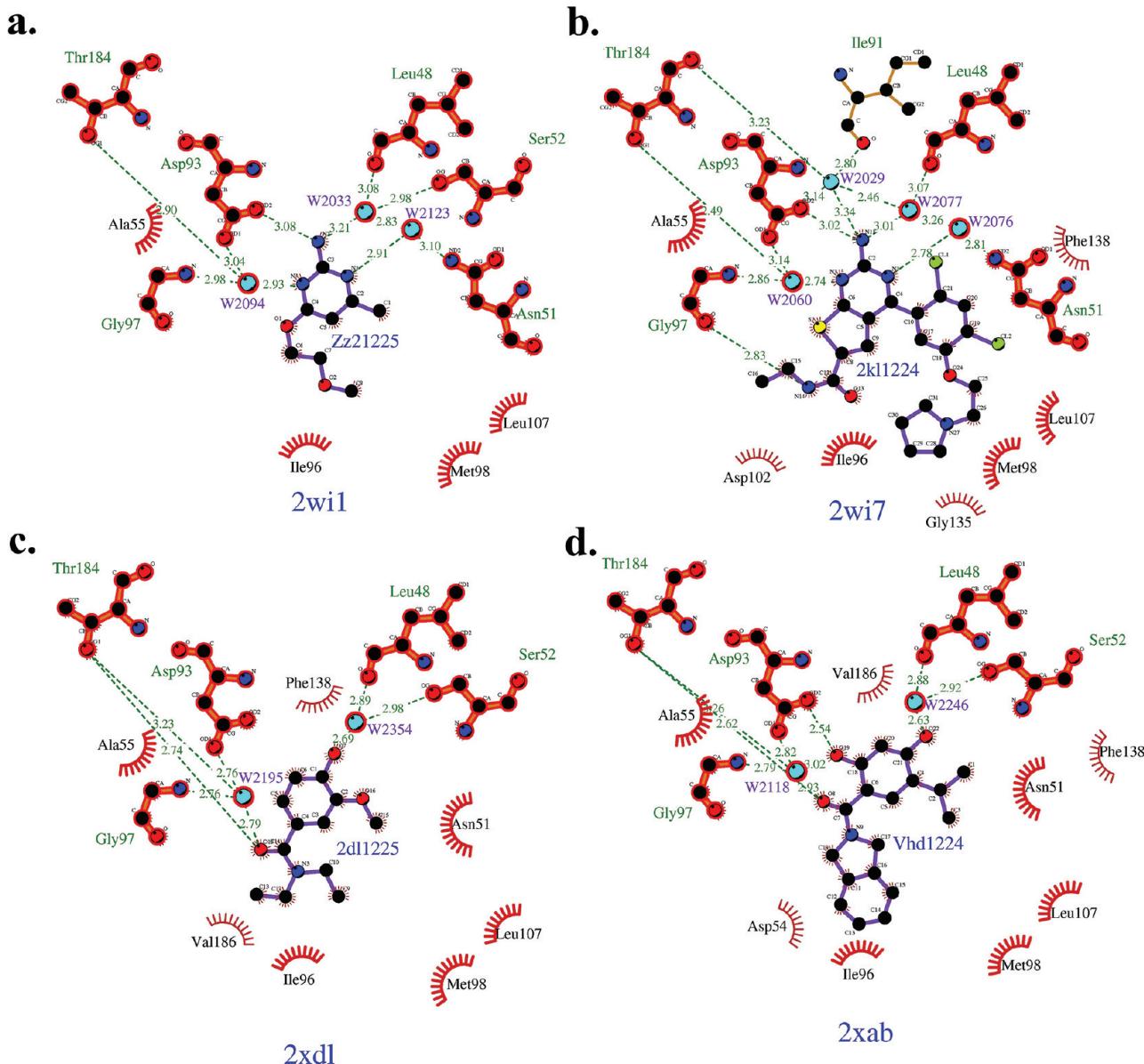
The two partial sequences are aligned to identify potentially equivalent residues in the two structures. If there are at least three pairs of equivalent residues, they are used for performing a least-squares superposition of the two 3D structures. After superposition, the root-mean-square deviation (rmsd) of the equivalenced residues is calculated. Any that have a rmsd > 3.5 Å

are deemed not equivalent, and the superposition is repeated without them.

If the resultant structural superposition is not reliable (i.e., fewer than three residues superpose with a rmsd of less than 2.5 Å), the program tries to fit on the ligands instead. It first uses subgraph matching to identify any common substructures, and then, if it has at least six equivalenced atoms, it fits the 3D structures based on these.

For very distantly related proteins and very dissimilar ligands, even this strategy may fail. So the program offers the option for the user to supply a structure-based alignment between the two or more proteins. Input formats include the CORA format from the CATHEDRAL structural alignment program<sup>15</sup> or a multiple alignment in FASTA format.

Once the 3D structures have been superposed, the program can identify which residues are in equivalent positions in 3D and which ligand atoms are in corresponding 3D locations. These equivalences



**Figure 5.** Interactions between different fragments bound to HSP90 in structures solved by two independent groups from Vernalis and Astex. All plots are displayed in a common orientation. Vernalis structures: (a) 2wi1 and (b) 2wi7; and Astex structures: (c) 2xdl and (d) 2xab.

are then used to guide the generation of the second LIGPLOT diagram. The 2D positions of the atoms in the first plot are used to constrain and restrain the 2D positions of the equivalent atoms in the second plot.

**Overlaying LigPlot+ Diagrams.** The first step is to flatten the second ligand into 2D and then translate, rotate, and flip it about its rotatable bonds to obtain the best fit to the flattened version of the first ligand. The goal is to minimize the rmsd between the equivalent atoms in the two flattened ligands. Often this gives a perfect fit; other times it fails, e.g., where the second ligand cannot make the right contortions to achieve the fit or one part has to fold back onto itself, resulting in its atoms clashing. In the latter case, the fit is sacrificed in order to avert atom clashes, and a warning is given to the user.

Where the ligands are of noticeably different sizes, the user will get the best results by plotting the complex with the largest ligand

first. If the second ligand is larger than the first, or extends beyond it in any direction, there is the possibility that it may overlap one or more of the protein residues on the first plot when the two plots are eventually overlaid. In this case, the program tries to push the relevant residues out of the way to give it enough space to prevent cluttering up the overlaid plots.

With the 2D conformation of the second ligand established, its interacting residues are added, in 2D, and their locations and orientations optimized by LigPlot+. Any equivalenced residues are restrained to the 2D locations and orientations of their partners in the first plot. The space taken by the first plot's ligand is designated as already occupied, so that, again, residues from one plot do not overlap the ligand in the other.

**Bond Order Information.** LigPlot+ can plot small molecules with the appropriate bond order information, provided it has been uploaded from an sdf or cif file.

**Problems of 2D Diagrams.** The depiction of protein–ligand interactions in 2D can greatly help one to understand them and see which residues are involved. However, there are limitations which should be borne in mind. First, some ligands just cannot be flattened into 2D, particularly those with multiply connected rings. Second, when a ligand is being flattened, subject to restraints from a prior plot, the program’s efforts to fit as many atoms from the two as it can sometimes leads to contortions in the ligand that need to be sorted out using the GUI. It is also important that the ligand molecule is correctly defined, either in a cif or sdf file, so that the supporting HBPLUS program can calculate all potential hydrogen bonds between ligand and protein.

## ■ RESULTS AND DISCUSSION

In this section we give some examples of LigPlot+ outputs. They are taken from two families that have been targets for extensive drug development: the protein kinases, for which a large amount of ligand–complex structural data is now deposited in the PDB, and the heat shock protein 90 (HSP90) superfamily which also has a range of data available, including for fragment-based structure design procedures.

**Different Compounds Binding to the Same Protein.** Figure 1 shows four plots of different ligands binding to the same protein, c-Abl kinase. In Figure 1a and b are two launched c-Abl inhibitors, imatinib (Gleevec) and nilotinib (Tasigna), PDB codes 2hy and 3cs9, respectively. Differences between the two ligands are confined to one part of the structure. The diagram emphasizes the high degree of similarity between the two binding modes, with equivalent residues circled in red, even though there are compensatory movements by corresponding secondary structure elements.<sup>16</sup> Although diagrams can be left in their automatically generated view, we elected to also make a small adjustment using the GUI for the ligand in Figure 1b, so that the hydrogen bonds to residues Glu 286 and Asp 381 were laid out equivalently; LigPlot+ had twisted the ligand to fit as many 3D-equivalenced atoms as it could in 2D. Also, a modification was required to the PDB’s Het Group Dictionary for imatinib so that the terminal nitrogen is protonated. With this adjustment the program identifies the potential hydrogen bonds to the carbonyl oxygens of Ile360 and His361 and includes them in the plot. This detail emphasizes issues that can occur when using public data.

Figure 1c and d shows two further c-Abl inhibitors from Parke-Davis; PD173955 (PDB code 1m52) and PD180970 (PDB code 2hzi), respectively. While the protein sequence remains identical, there is some structural rearrangement on binding these ligands. Because of the structural rearrangement, Phe382 is now located in the 3D position previously occupied by Asp381 when binding the imatinib and nilotinib. The LigPlot+ figures reflect this; circles highlight equivalent positions in 3D but allow the residues to change where necessary. So, Phe382, in Figure 1c and d is now plotted in the location that the Asp381 previously occupied in Figure 1a and b.

**Same Compound Binding to Homologous Protein Targets.** Figure 2 gives an example of the same compound binding to different, but closely related, proteins. It shows imatinib bound to both c-Abl (2hy) and c-Kit (1t46) kinases; it is now known that in both cases imatinib has a therapeutic effect. Of importance to the researcher are the structurally equivalent positions with similar binding roles. These are clearly and immediately identified in the plot, even when the amino acid types vary. For instance, it is a main-chain interaction between the ligand and residue Met318 (2hy)/

Cys673 (1t46) that is of importance, so the side-chain methionine/cysteine substitution is of only limited relevance to ligand binding. The sequence identity of the two proteins’ tyrosine kinase domains, in which their binding sites are located, is 39% overall, but it is much higher for the binding site residues. For example, of the 18 equivalenced (and highlighted) residues shown in the plot, 13 (72%) are of identical types. Of course, also of interest are the differences between binding sites, as these provide the opportunity for design of selective inhibitors. These, too, can easily be identified.

**Similar Ligands Binding to Homologous Proteins.** Comparative diagrams can also be generated when both the ligand and the protein differ. In Figure 3 we compare the previous c-Abl/imatinib (2hy) complex with a motesanib/VECFR2 kinase complex published by Amgen Inc. (3efl). Here the protein sequence identity is approximately 35%, whereas the Tanimoto coefficient for the ligands is 0.9 (calculated by University of Padova server).<sup>17</sup> But higher binding site conservation, relative to the overall similarity between the two proteins, is quickly communicated to the user by the diagrams, and thus the rationale for both proteins binding similar compounds can be quickly appreciated.

**Binding Site Similarities in Very Distant Homologues.** It is now well established that protein structure and function can remain conserved, even in the absence of significant sequence identity. Such relationships are elegantly described through structure classifications, such as CATH<sup>18</sup> and SCOP.<sup>19</sup> When the functional sites of distant homologues bind similar or identical small molecules, such as nucleoside phosphates, a comparison of those can be extremely productive, both for drug discovery as well as basic research.

To illustrate, we use the ATP binding sites of three distantly related proteins: DNA gyrase B, which is the target of novobiocin, a marketed antibiotic derived from *Streptomyces niveus*; heat shock protein 90 (HSP90) an anticancer target with several compounds in clinical trials; and osmolarity sensing histidine kinase, a potential target for a next-generation antimicrobial.

Overall, the ATP binding domains have considerably less than the ~30% identity required for alignment tools to detect a homologous relationship based on sequence alone, yet structures are clearly conserved.<sup>20</sup> Thus, a structure-based sequence alignment must be generated to determine the equivalent residue positions in each protein. Here we use CATHEDRAL,<sup>15</sup> which can objectively align multiple structures without human intervention. Other methods with similar objectives are also available.<sup>21</sup> From the resulting alignment (Figure 4a), it is apparent that globally, the sequences have very little (~15%) identity with one another.

Figure 4b shows the corresponding diagram generated by LigPlot+ from the CATHEDRAL alignment. Taking histidine kinase (PDB code 3d36) as an arbitrary reference, we observe that for DNA gyrase B (1z5a) there are 18 structurally equivalent positions participating in ATP binding (some via water molecules), 8 of which (44%) have identical residues in each structure. Similarly, looking at HSP90 (1byq), there are 14 structurally conserved positions of which 8 (57%) have identical residues. Furthermore, there are six structurally conserved interactions that have the same amino acid in each of the three therapeutically relevant proteins. Of particular prominence is an aspartate residue that consistently hydrogen bonds to the amine group of the adenosine. If conservative substitutions, such as Ser/Thr, as well as structurally conserved positions that only require main-chain interactions are also considered, binding site conservation is even more striking.

**Application to Fragment-Based Drug Design.** Over the past decade, fragment-based drug design has developed from a concept to a mainstream approach. Results for fragments designed to bind

HSP90 (one of the three proteins referred to in the previous section) have been reported by two separate groups from Vernalis and Astex.<sup>22,23</sup> One key challenge with such fragments is the large number of structures that must be analyzed and compared in an objective, time-efficient manner. This is a key objective of LigPlot+, which we illustrate with reference to these publications.

In the Vernalis paper,<sup>22</sup> a series of more than 50 fragments are described. From these, seven crystallographic structures tracking the development of the series were made available. The earliest fragment of the series with a deposited structure (PDB code 2wi1) highlights use of the aspartate NH<sub>2</sub>—adenosine interaction, conserved between histidine kinase, DNA gyrase, and HSP90 (Figure 5a). In fact, this interaction (Asp93 in HSP90) persists through all the subsequent compounds to the final complex (PDB code 2wi7, Figure 5b).

A similar situation is observed in the subsequent paper from Astex.<sup>23</sup> The deposited complexes for compounds 3 (2xdl, Figure 5c) as well as 21 and 24 (2xht and 2xhx, not shown) hydrogen bond through an intermediate water to Asp93. But by compound 31 (2xab, Figure 5d), this has become a direct hydrogen-bond interaction. Other characteristics referred to in the publication are also apparent from the LigPlot+ diagrams, such as the movement of Lys58, which provides a hydrogen bond for compound 21 but is subsequently displaced by the introduction of an isoindoline.

LigPlot+ is also able to highlight relevant water molecules which, if displaced, could accommodate further changes to the compound. It is therefore of practical interest to note that the complex for compound 31 (2xab, Figure 5d) and the resultant development candidate AT13387 (2xjx, data not shown) retain water molecules in the binding site.

Thus, where available, the display of water molecules around the ligand binding site can inform new design strategies, by pinpointing voids that could be occupied by modifications to the ligand.

## CONCLUSION

We have introduced a significant reworking of the well-known LIGPLOT software. By enabling the direct comparison of multiple, related protein complexes, while allowing for the simultaneous variation of both protein sequence and compound formula, LigPlot+ has evolved into a valuable investigative tool. We anticipate that the approaches described here will be of value to both domain experts (such as crystallographers and modellers) as well as those involved in compound design and synthesis. Such tools should also facilitate the wider interpretation of potentially high-value structural information in research.

## AUTHOR INFORMATION

### Corresponding Author

\*E-mail: roman@ebi.ac.uk

## ACKNOWLEDGMENT

LigPlot+ is available from <http://www.ebi.ac.uk/thornton-srv/software/LigPlus>.

## REFERENCES

- Wallace, A. C.; Laskowski, R. A.; Thornton, J. M. LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein Eng.* **1995**, *8*, 127–134.

- McDonald, I. K.; Thornton, J. M. Satisfying hydrogen bonding potential in proteins. *J. Mol. Biol.* **1994**, *238*, 777–793.
- Laskowski, R. A. PDBsum new things. *Nucleic Acids Res.* **2009**, *37*, D355–359.
- Berman, H. M.; Battistuz, T.; Bhat, T. N.; Bluhm, W. F.; Bourne, P. E.; Burkhardt, K.; Feng, Z.; Gilliland, G. L.; Iype, L.; Jain, S.; Fagan, P.; Marvin, J.; Padilla, D.; Ravichandran, V.; Schneider, B.; Thanki, N.; Weissig, H.; Westbrook, J. D.; Zardecki, C. The Protein Data Bank. *Acta Crystallogr., D: Biol. Crystallogr.* **2002**, *58*, 899–907.
- Stierand, K.; Maass, P. C.; Rarey, M. Molecular complexes at a glance: automated generation of two-dimensional complex diagrams. *Bioinformatics* **2006**, *22*, 1710–1716.
- Fricker, P. C.; Gastreich, M.; Rarey, M. Automated drawing of structural molecular formulas under constraints. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1065–1078.
- Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. A fast flexible docking method using an incremental construction algorithm. *J. Mol. Biol.* **1996**, *261*, 470–489.
- Clark, A. M.; Labute, P. 2D depiction of protein-ligand complexes. *J. Chem. Inf. Model.* **2007**, *47*, 1933–1944.
- Schrödinger, LLC. *Ask the Scripts Expert*; <http://www.schrodinger.com/newsletter/18/118/>. Accessed August 1, 2011.
- Sayle, R. A.; Milner-White, E. J. RASMOL: biomolecular graphics for all. *Trends Biochem. Sci.* **1995**, *20*, 374.
- Schrödinger, LLC. The PyMOL Molecular Graphics System, Version 1.2r3pre.
- Berman, H.; Henrick, K.; Nakamura, H.; Markley, J. L. The worldwide Protein Data Bank (wwPDB): ensuring a single, uniform archive of PDB data. *Nucleic Acids Res.* **2007**, *35*, D301–303.
- wwPDB. *File format documentation*; <http://www.wwpdb.org/docs.html>. Accessed August 1, 2011.
- Needleman, S. B.; Wunsch, C. D. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J. Mol. Biol.* **1970**, *48*, 443–453.
- Redfern, O. C.; Harrison, A.; Dallman, T.; Pearl, F. M.; Orengo, C. A. CATHEDRAL: a fast and effective algorithm to predict folds and domain boundaries from multidomain protein structures. *PLoS Comput. Biol.* **2007**, *3*, e232.
- Weisberg, E.; Manley, P. W.; Breitenstein, W.; Bruggen, J.; Cowan-Jacob, S. W.; Ray, A.; Huntly, B.; Fabbro, D.; Fendrich, G.; Hall-Meyers, E.; Kung, A. L.; Mestan, J.; Daley, G. Q.; Callahan, L.; Catley, L.; Cavazza, C.; Azam, M.; Neuberg, D.; Wright, R. D.; Gilliland, D. G.; Griffin, J. D. Characterization of AMN107, a selective inhibitor of native and mutant Bcr-Abl. *Cancer Cell* **2005**, *7*, 129–141.
- University of Padova. *Molecular Modeling Section*; Molecular Modeling Section Lab: Padova, Italy; <http://mms.dsfarm.unipd.it/>. Accessed August 1, 2011.
- Cuff, A. L.; Sillitoe, I.; Lewis, T.; Clegg, A. B.; Rentzsch, R.; Furnham, N.; Pellegrini-Calace, M.; Jones, D.; Thornton, J.; Orengo, C. A. Extending CATH: increasing coverage of the protein structure universe and linking structure with function. *Nucleic Acids Res.* **2011**, *39*, D420–426.
- Andreeva, A.; Howorth, D.; Chandonia, J. M.; Brenner, S. E.; Hubbard, T. J.; Chothia, C.; Murzin, A. G. Data growth and its impact on the SCOP database: new developments. *Nucleic Acids Res.* **2008**, *36*, D419–425.
- Tanaka, T.; Saha, S. K.; Tomomori, C.; Ishima, R.; Liu, D.; Tong, K. I.; Park, H.; Dutta, R.; Qin, L.; Swindells, M. B.; Yamazaki, T.; Ono, A. M.; Kainosh, M.; Inouye, M.; Ikura, M. NMR structure of the histidine kinase domain of the *E. coli* osmosensor EnvZ. *Nature* **1998**, *396*, 88–92.
- Hasegawa, H.; Holm, L. Advances and pitfalls of protein structural alignment. *Curr. Opin. Struct. Biol.* **2009**, *19*, 341–348.
- Brough, P. A.; Barril, X.; Borgognoni, J.; Chene, P.; Davies, N. G.; Davis, B.; Drysdale, M. J.; Dymock, B.; Eccles, S. A.; Garcia-Echeverria, C.; Fromont, C.; Hayes, A.; Hubbard, R. E.; Jordan, A. M.; Jensen, M. R.; Massey, A.; Merrett, A.; Padfield, A.; Parsons, R.; Radimerski, T.; Raynaud, F. I.; Robertson, A.; Roughley, S. D.; Schoepfer,

J.; Simmonite, H.; Sharp, S. Y.; Surgenor, A.; Valenti, M.; Walls, S.; Webb, P.; Wood, M.; Workman, P.; Wright, L. Combining hit identification strategies: fragment-based and in silico approaches to orally active 2-aminothieno[2,3-d]pyrimidine inhibitors of the Hsp90 molecular chaperone. *J. Med. Chem.* **2009**, *52*, 4794–4809.

(23) Murray, C. W.; Carr, M. G.; Callaghan, O.; Chessari, G.; Congreve, M.; Cowan, S.; Coyle, J. E.; Downham, R.; Figueroa, E.; Frederickson, M.; Graham, B.; McMenamin, R.; O'Brien, M. A.; Patel, S.; Phillips, T. R.; Williams, G.; Woodhead, A. J.; Woolford, A. J. Fragment-based drug discovery applied to Hsp90. Discovery of two lead series with high ligand efficiency. *J. Med. Chem.* **2010**, *53*, 5942–5955.