JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

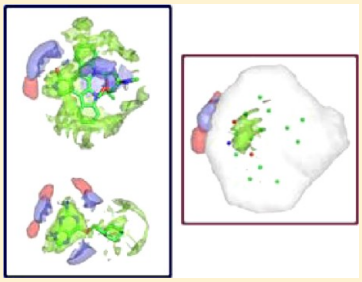
GRID-Based Three-Dimensional Pharmacophores I: FLAPpharm, a Novel Approach for Pharmacophore Elucidation

Simon Cross,*[,†] Massimo Baroni,[†] Laura Goracci,[‡] and Gabriele Cruciani[‡]

[†]Molecular Discovery Limited, 215 Marsh Road, Pinner, Middlesex, London HA5 5NE, United Kingdom

[‡]Laboratory for Chemometrics and Cheminformatics, Chemistry Department, University of Perugia, Via Elce di sotto 10, I-06123 Perugia, Italy

ABSTRACT: Pharmacophore elucidation approaches are routinely used in drug discovery, primarily with the aim of determining the three-dimensional arrangement of common features shared by ligands interacting at the site of interest; these features can then be used to investigate the structure−activity relationship between the ligands and also to screen for other molecules possessing the relevant features. Here we present a novel approach based on GRID molecular interaction fields and the derivative method FLAP that has been previously described, which provides a common reference framework to compare both small molecule ligands and macromolecular protein targets. Unlike classical pharmacophore elucidation approaches that extract simplistic molecular features, determine those which are common across the data set, and use these features to align the structures, FLAPpharm first aligns the structures and subsequently extracts the common interacting features in terms of their molecular interaction fields, pseudofields, and atomic points, representing the common pharmacophore as a more comprehensive pharmacophoric pseudomolecule. The approach is applied to a number of data sets to investigate performance in terms of reproducing the X-ray crystallography-based alignment, in terms of its discriminatory ability when applied to virtual screening and also to illustrate its ability to explain alternative binding modes. In part two of this publication, a comprehensive benchmark data set for pharmacophore elucidation is presented and the performance of FLAPpharm discussed.



INTRODUCTION

Understanding how molecules interact with a therapeutic target of interest is clearly of great importance in drug discovery, and the explosion of X-ray crystallographic information in the protein databank reinforces the importance of this need; once the three-dimensional structure of a target cocrystallized with a bound ligand is available, it immediately offers ideas about which parts of the chemistry of the ligand are interacting, which can be modified to modulate other parameters (such as pharmacokinetic properties or intellectual property), and potentially where new interactions with the target can be made. In the absence of experimental data describing the ligand of interest and the target, computational approaches such as docking can be used to predict how the molecule might interact with the receptor. If no crystallographic structure is available of the target, ligand-based approaches such as pharmacophore elucidation can be used to predict the important chemical features that are required for biological activity, given a handful of ligands that interact at the same site. Even if structure-based approaches can be applied, it is common practice to use alternative methods to reinforce the available information; this makes the ligand-based approaches some of those most widely used in computational drug discovery and, therefore, critical in supporting the discovery process.

The term pharmacophore was introduced over forty years ago[1] and the in silico prediction of the ligand binding conformation from a set of active structures followed shortly afterward.[2] The pharmacophore of a protein target typically refers to the spatial arrangement of a small set of features shared by compounds that interact at the site of interest; the features being abstract terms relating to molecular interactions (hydrogen-bond donor, hydrogen-bond acceptor, hydrophobic, and charged groups). These features are required but not necessarily sufficient for molecular recognition. It is also important to consider that multiple binding modes for a particular site may exist, and searching for common pharmacophoric features between molecules that exhibit different binding modes is inherently flawed.

Usually, the pharmacophore features are defined using a set of rules relating the functional group chemistry to the feature type, and for hydrogen-bond donors and acceptors "extension points" representing the complementary interaction site are also incorporated. A detailed review of pharmacophore methods used in discovery has recently been published;[3] the typical applications include the elucidation of the pharmacophore required for a binding interaction and the screening of molecules to determine whether or not they can match the pharmacophore. Given that the features are defined according to a set of rules, one potential problem is that the features are defined adequately.[4] Typically these features are extracted from a conformational ensemble of each molecule under consideration and molecules overlaid by least-squares fitting of the common feature points. The common points themselves are

Received: March 22, 2012
Published: September 12, 2012

   dx.doi.org/10.1021/ci300153d | *J. Chem. Inf. Model.* 2012, 52, 2587−2598

usually found by methods such as clique-detection, and typically a number of three of four point pharmacophores models are generated. A root mean squared deviation (rmsd) threshold is used to determine if the configuration of features is common, and some approaches will allow a specified number of molecules to miss a certain feature. Another potential problem, then, is that these parameters can be adjusted, introducing bias into the approach. It may also occur that there are many potential common configurations of features, resulting in more time spent on analyzing the results. The models are also typically scored according to a function that takes into account aspects such as the feature matching, volume overlap, and strain energy of the model structures, and these can also be weighted, allowing more flexibility as well as user bias.

In our experience, a problem with this approach in general is that the identified common features (typically three of four points) may be well-aligned but the molecules themselves are not; partial matching is heavily dependent on user bias, and some molecules may not be aligned at all. We therefore hypothesized that a better approach might be to align the molecules first and, subsequently, extract the common features from this alignment. GRID molecular interaction fields (MIFs)[5] are extremely useful to describe molecules in terms of how they interact with their environment, and a wide range of applications to drug discovery have previously been reported[6−15] (reviewed in refs 16 and 17). The use of GRID MIFs for molecular alignment in the pharmacophore similarity program FLAP has also been previously reported,[18] whereby the alignment is scored according to GRID MIFs and can therefore be used for virtual screening. Both prospective[19−21] and retrospective[22] virtual screening studies have been reported, illustrating the performance of FLAP; therefore, we decided to adapt this approach to pharmacophore elucidation. In the following sections, we describe the approach, the performance in terms of reproducing experimentally three-dimensional molecular alignments, and also in terms of virtual screening.

Pharmacophore Elucidation Using GRID Molecular Interaction Fields. Our ultimate aim was to produce a method that could take as input two-dimensional structures of active molecules that bind at the same site and, with as little user bias as possible, reproduce the experimentally derived alignments given by the same structures (cocrystallized with their targets, protein targets aligned according to their backbone atoms, ligands then extracted) and in a reasonable time frame. From these alignments, the common features can be extracted in a straightforward manner (described in more detail below).

To achieve this, several questions needed addressing:

1. How should we validate the approach?
2. How should we extract the common pharmacophore?
3. Can we reproduce the X-ray binding conformation from a two-dimensional structure?
4. Given the X-ray conformations can we reproduce the correct alignment?
5. Given a set of possible alignments, can we select the correct model as the best one?
6. Can we avoid depending on an arbitrary template?

For the initial validation of the method, we decided to use the data set published by Patel and co-workers,[23] since it was a comparison of several methods and, for five data sets, contained several diverse chemotypes. In the results section we will discuss the performance of FLAPpharm using these data sets, in addition to other examples. A secondary goal is that of virtual screening performance, for which we revisited the DUD data set,[24] building models using the DUD chemotype cluster parents as identified by Good[25] and, then, using the top scoring models to screen each of the thirteen data sets that we discussed previously.[22] In part II of this publication, we have assembled a benchmark data set for pharmacophore elucidation, consisting of 81 targets containing 960 ligands,[26] and investigate the performance of FLAPpharm.

Extracting the Common Pharmacophore. As described above, rule-based features, in addition to their simplistic nature, suffer from some limitations. A more detailed description of molecules and how they appear to their interacting receptors is provided by molecular interaction fields. In FLAP, typically the GRID probes H, O, N1, and DRY are used to describe shape, hydrogen bond acceptor, hydrogen bond donor, and hydrophobic interactions. Charge is implicitly described by the magnitude of the interaction using the N1 and O probes (for more accurate prediction of the electrostatics the probes O::, N+, or N1+ can also be used, although the FLAP field similarity does not usually require this improved electrostatic accuracy). The common molecular interaction fields (MIFs) can be found simply by taking the mean field values across the (aligned) data set at each position on the grid; interactions common to all N molecules will remain at a similar interaction energy, interactions common to only one molecule will be reduced to $1/N$th the original intensity. Unlike rule-based approaches, partial matching is inherently described; interactions common to 3 out of 6 aligned molecules will be reduced to approximately 50% of their original intensity; this partial matching is clearly visualized by the intensity of the MIF contours. Henceforth, we are introducing a new term *pharmacophoric interaction fields (PIFs)* to describe these common molecular interaction fields.

FLAP also uses pseudofields (or pseudoMIFs), which can be thought of as electron-density like fields centered on each atom, which correspond to the chosen probe types (for FLAPpharm the O, N1, and DRY probes). An analogous grid cage is used, however rather than measuring the interaction energies at each grid point; the proximity to the molecule's corresponding feature-type is described by a pseudopotential that has been calibrated such that complete overlap by a molecular feature in the corresponding receptor MIF would generate the maximum field similarity at that position. For example, the O MIF describing an acceptor interaction with a protein might be expected to overlap with the O (acceptor) pseudofield of a ligand carbonyl oxygen atom and that atom's N1 (donor) MIF would likewise interact with the equivalent protein N1 (donor) pseudofield. In the same manner as for the MIFs, the mean pseudofield values can be used to represent the common atomic positions, and as above, we will term these *pseudoPIFs*. The centroid of these common atomic positions are also retained as pharmacophoric points. With this simple approach, a set of aligned molecules is converted from the individual molecule atoms, MIFs, and pseudoMIFs, to a *pharmacophoric pseudomolecule* (which we will abbreviate as the pharmacophore) consisting of *pharmacophoric points*, *PIFs*, and *pseudoPIFs*. The same common reference framework is therefore available for the pharmacophore as for individual molecules. One advantage of this is that a FLAPpharm model can therefore be used in FLAP just like any other molecule; for example, it could be docked into a receptor which could

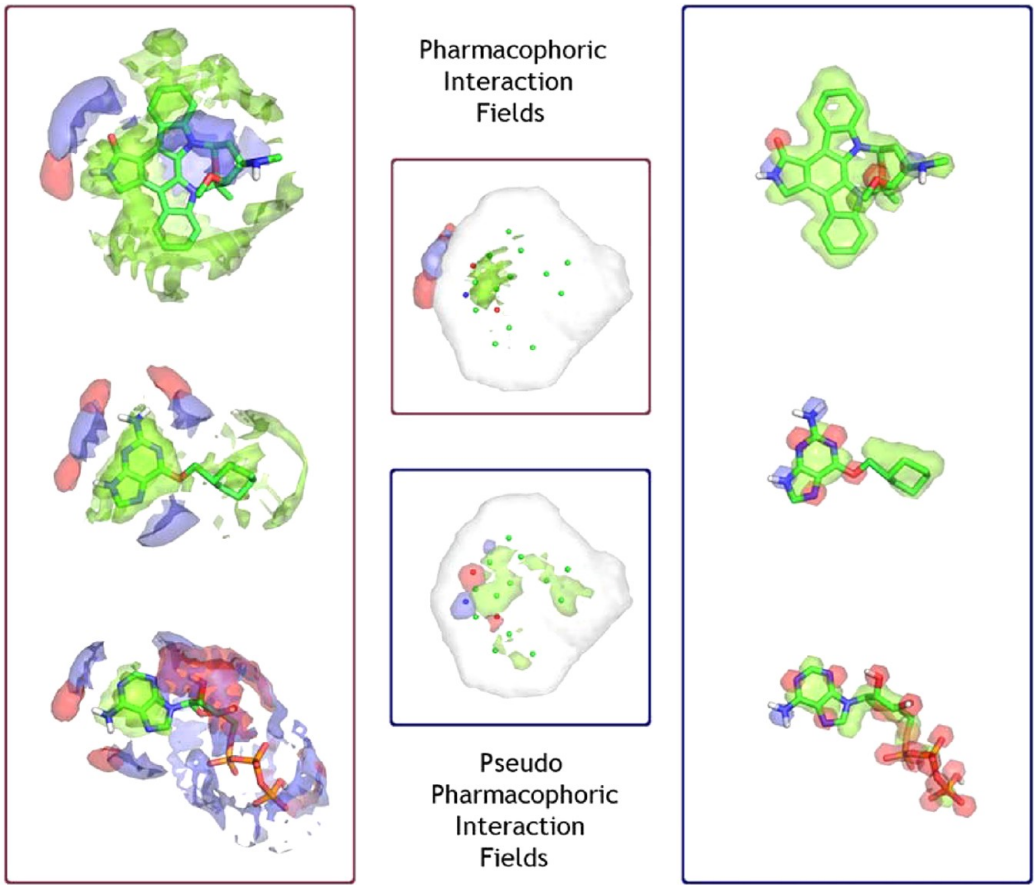


Figure 1. GRID molecular interaction fields (left) and pseudofields (right) for six CDK2 ligands (three shown for clarity) converted to a pharmacophoric pseudomolecule (middle) consisting of the common donor/acceptor/hydrophobic atom locations, pharmacophoric interaction fields (PIFs), and pharmacophoric pseudofields (pseudoPIFs). The data set reduces to a common hydrophobic patch above and below the plane of the aromatic ring system (green contour, top middle), a donor/acceptor interaction pair (blue/red), and a partial acceptor interaction (smaller red contour, top of picture). The corresponding donor/acceptor/hydrophobic pharmacophoric pseudofields (pseudoPIFs) are also shown (botton middle, blue/red/green contours). The hydrophobic pseudoPIFs on the right-hand side of the image have a weaker PIF interaction which is therefore not shown in the PIF isocontour.

reinforce or disprove the hypothesis. Figure 1 illustrates the FLAPpharm pharmacophoric representation for six crystallographic CDK2 ligands.

Reproducing the Crystallographic Ligand Conformation. In order to be able to reproduce the crystallographic derived alignments, the method clearly requires a conformer generator that is able to reproduce the binding conformation. It was also important to us to consider performance, our aim therefore was to be able to produce the relevant conformation within a pool of the smallest possible number of generated conformers. The conformer generator within FLAPpharm is based on the MM3 forcefield;[27] a random search is performed using 40 samples for each rotatable bond, which produces thousands of potential conformers. These are then minimized, and conformers are removed if their energy is greater than 20 kcal/mol above the minimum. The conformers are then ranked by energy and filtered using a threshold rmsd to another lower energy conformer. The user can then specify the number of available lowest energy conformers to consider in the pharmacophore elucidation step; 30 is usually satisfactory and even using as low as 10 conformers per molecule can give excellent results. To validate the conformer generator, we generated 30 conformations for each molecule under consideration, aligned the conformation to its equivalent X-ray structure, and calculated the rmsd.

For the Patel data set[23] of 35 ligands (5 targets), all but 5 were reproduced with an rmsd < 2.0 Å. Of these 5, 2 contain 10 rotatable bonds (1OHK, 4TMN) and are therefore very flexible molecules, 1RT5 contains a *cis*-thiamide which is an unexpected conformation and also not found in the Cambridge Structural Database (CSD),[28] 1D6W contains a boat cyclohexyl group that is also not found in the CSD, and 1KLM contains an equatorial pyridyl-piperidine which is entirely reasonable, although the axial conformation is also possible according to the CSD and depends on the ring substitution. For the remainder of this study, these five problematic ligands were removed. We then expanded the validation to a larger set, using the PDBbind[29] refined data set of 1300 molecules. We filtered this data set according to drug-likeness[30] to leave 730 ligands from high quality crystallographic structures. For this set, 91% were reproduced at <2.0 Å, and we were satisfied that our conformer generation methodology was good enough to move forward.

Reproducing the Correct Alignment Using X-ray Conformations. As discussed above, we have already demonstrated that FLAP gives excellent results in terms of discriminating actives and decoys in virtual screening, after

dx.doi.org/10.1021/ci300153d | J. Chem. Inf. Model. 2012, 52, 2587−2598

Table 1. Patel Data Set Ligands, Aligned to Its Best Performing Template Molecule, with the rmsd Compared to Its X-ray Alignment

target	PDB	template	rmsd	target	PDB	template	rmsd
cdk2	1AQ1	1FVV	0.54	thrombin	1C4V	1D9I	0.53
	1E1X	1AQ1	1.20		1D4P	1DWD	0.15
	1FVV	1AQ1	1.69		1D6W	1D4P	0.58
	1DI8	1AQ1	0.67		1D9I	1C4V	0.47
	1E1V	1AQ1	0.92		1DWD	1D4P	0.16
	1FIN	1E1X	2.49		1TOM	1DWD	0.27
					1FPC	1TOM	1.59
dhfr	1BOZ	1HFP	0.50				
	1DLR	1HFP	1.09	hivrt	1BQM	1FK9	1.02
	2DHF	1DLR	0.71		1DTT	1FK9	1.21
	1OHK	1BOZ	0.19		1EP4	1RT1	1.91
	1HFP	1BOZ	0.58		1FK9	1VRU	1.48
	1DRF	1HFP	0.44		1RT1	1FK9	1.55
					1RT3	1VRU	0.62
thermolysin	4TMN	1QF1	0.62		1VRU	1FK9	1.02
	1HYT	7TLN	0.76		1TVR	1FK9	0.75
	1QF1	5TMN	0.77		1RT5	1FK9	1.00
	5TLN	7TLN	0.78		1KLM	1RT1	9.50
	5TMN	1QF1	0.98				
	7TLN	5TLN	0.69				

aligning to a query template (or receptor pocket).[22] However, we wanted to prove that the alignments were correct and that the discrimination was not giving excellent results for the wrong reasons. To test the alignment, we retained the crystallographic conformations and, for each data set, aligned each ligand to each of the others in their X-ray position. The aligned ligand was then compared to its X-ray position, and the rmsd was calculated. The best results are shown in Table 1.

Almost all molecules contain a template that is able to align them correctly in order to reproduce their X-ray derived alignment; for 1FIN the alignment is correct except for the triphosphate tail. FLAP is therefore able to align the molecules, given an appropriate template, to reproduce their experimentally derived alignment. It is also worth noting that this experimentally derived alignment is based on the protein backbone of each crystallographic structure, which is a source of potential error.

Inspection of these results also highlighted a potential problem: if A is used as a template to align B and C, both B and C might align reasonably according to the X-ray alignments; however, B might not necessarily align well to C. The best template to align each molecule is not always the same across the data set; therefore attempting to use only a single template to align the other structures is likely to give suboptimal results.

Generating the Common Alignment and Extracting the Common Pharmacophore: FLAPpharm. The first step in the procedure is to prepare the input molecules. As described above, a number of low energy conformers (typically up to 30) is generated for each molecule (we have shown above that this sampling is enough to include a conformation close to the bioactive one in ~90% of cases). Of course, the user can also provide the X-ray conformations as input, to focus the method on the known bioactive conformations or, alternatively, can use conformers generated elsewhere as input to the program.

For each of these conformations, the molecular interaction fields for the H, O, N1, and DRY probe are calculated at a 0.75 Å grid resolution. The grid cage itself is aligned to the principal axes of the molecule, to ensure consistent results regardless of any provided input coordinates. Once the MIFs are available, the FLAP fingerprints are generated for each conformation. These fingerprints provide two alternatives: the fingerprints can be directly compared to provide a FLAP similarity score, or they can be used to align a structure to a template using all combinations of the quadruplets. In this case, the MIF similarity is calculated. The first approach is of course much faster, the second is both more accurate and provides the necessary alignment.

The next step is a pairwise comparison of all molecules and their conformations using the fast FLAP pharmacophoric fingerprints. Each conformation of each molecule can then be scored across all combinations of conformations from the other molecules in the data set, and the best combination score is retained. The top N conformations for each molecule are then retained, representing those conformers that are most pharmacophorically similar to the conformations in the rest of the data set. In fast mode, $N = 5$; on other modes, it is 10. Ultimately, this step is yet another step in the conformer generation cascade, where fairly comprehensive conformational sampling is sequentially filtered by energy, diversity, and finally pharmacophoric relevance depending on the data set.

The final step covers the main search algorithm. Let A_c1 represent conformer 1 of molecule A and be our primary template molecule conformer, which is the first molecule added to the model. Then all conformers of B (B_c1 ... B_c10) are aligned to this template conformer. The best scoring conformer of B (e.g., B_c3) is then selected and added to the model. All conformers of molecule C are then aligned both to A_c1 and B_c3. The model score in this case is the mean of the pairwise similarities of these aligned conformers of AB, AC, and BC.

This process is then repeated for all other molecules in the data set. It is important to note that this search is repeated using each conformer of each molecule as the initial template molecule. In this way we avoid the problem of selecting an arbitrary template: at some point each molecule will be aligned to its best template, and if a molecule is not appropriate for the

model (ie. it has a different binding mode), it should only have a minor impact on the final model.

The scoring function used is a weighted sum of the field similarities (shape, hydrophobic, donor, acceptor), with the shape similarity weighted at 50%. The rationale behind the shape similarity being downweighted is that a priori there is no knowledge of the shape of the binding pocket. Therefore, the interaction features must drive the alignments, and following this, the shape of the pseudocavity can be inferred (this contrasts with our experience using these fields in docking-like approaches, where the weight of the shape field is much more important). After this process is completed, many potential models are produced. Each model consists of these molecule alignments, and the common pharmacophoric pseudomolecule that is generated (with pharmacophoric points, PIFs, and pseudoPIFs). In our experience, a good model is usually found in the top 5 ranked models, and the results will be described below as well as in part II of this paper.[26] The pharmacophoric pseudomolecule can be subsequently used in FLAP as a template for virtual screening, to dock into a receptor pocket to help validate the hypothesis, and to aid SAR analysis by inspection of the field contours, and the molecular alignments can be used for three-dimensional (3D) quantitative structure activity relationship (QSAR). All of these applications are currently available in the FLAP program and will be the subject of future publications.

FLAPpharm Generated Alignment Models. FLAPpharm was run on each of the Patel data sets. To test how the alignment aspect of the algorithm performed, the X-ray conformations were used as input (rigid alignment). The starting coordinates of the structures were rotated and translated randomly to avoid starting from the optimal alignment solution. To test the full model building, the structures were converted to 2D before use, and the algorithm run using up to 30 generated conformations (unbiased model). The results for each of the data sets is analyzed below, comparing the ligand alignments and obtained pharmacophoric pseudomolecules for the X-ray derived alignments, FLAPpharm rigid alignments, and FLAPpharm flexible alignments. In each case only the top-ranked model is analyzed.

Cyclin-Dependent Kinase 2 (CDK2). The target pharmacophore as identified by Patel et al. consists of simply an acceptor and a hydrophobic region, giving rise to a donor field and hydrophobic field, respectively. Figure 2A illustrates the alignment of the ligands according to their protein-partner backbones, and the pharmacophoric interaction fields identified according to FLAPpharm. Given the nature of the pharmacophoric interaction field as described above, two additional donor interactions are found: one common to five of the six ligands (1DI8 does not contain it), and the other common to two of the ligands (1E1V, 1E1X, there is some ambiguity as to whether 1FVV also contains this feature). These ligands also illustrate why it is important to downweigh the shape during alignment; for example, 1FVV overlaps very little with 1E1V, 1E1X, and 1FIN.

Figure 2B illustrates FLAPpharm rigid alignment. The main features identified from the X-ray alignments are all found, and on closer inspection, the partial donor common to 1E1V and 1E1X now has 1FVV aligned so that it fits this feature more closely (hence increasing the magnitude of the associated fields); 1DI8 is incorrectly aligned but is also aligned to fit this feature. One advantage of the FLAPpharm pharmacophoric interaction field is that even though 1DI8 is aligned incorrectly

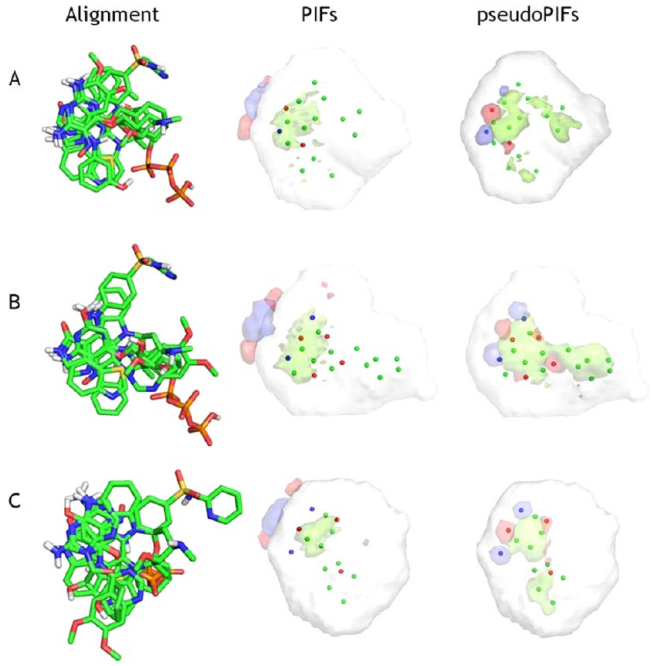


Figure 2. CDK2 ligand alignments and their respective pharmacophoric pseudomolecules. The pseudomolecules are shown twice, with either their pharmacophoric interaction fields (PIFs) or pharmacophoric pseudofields (pseudoPIFs) displayed, for clarity. The ligands are taken from (A) the X-ray-derived alignements, (B) the FLAPpharm rigid alignments, and (C) the FLAPpharm unbiased model alignments. Hydrophobic atoms and fields are displayed in green, hydrogen-bond donor atoms and fields are displayed in blue, hydrogen-bond acceptor atoms and fields are displayed in red, and the shape field is shown in white.

in general, the main donor/acceptor field pair is hardly affected at all. The algorithm has also aligned some acceptor atoms from several of the molecules, giving rise to two additional acceptor pseudoPIFs.

Figure 2C illustrates the top-ranked model when using the 2D molecules as input and generating conformations for each molecule. The PIFs are almost identical to those found by the X-ray alignments. Although the conformations for some of the molecules are different, the key interacting groups are all aligned correctly, with the exception of 1DI8, as before. In this case 1DI8 is aligned so that its phenolic hydroxyl group is positioned so that it can make its donor field overlap with the main donor interaction, and its acceptor field overlap with the partial acceptor at the top left of the displayed pharmacophore.

There are also some hydrophobic points and pseudoPIFs identified that differ from the X-ray and rigid alignments; however in these regions, the ligands only have weak hydrophobic MIFs; hence, there is no PIF identified. The reason that these regions are in different locations is likely due to the fact that there is no conformational restriction favoring the X-ray alignment over the unbiased model in this location and, also, that there is no PIF there to favor or disfavor one over the other. Starting from two-dimensional input, and reproducing close to the X-ray alignments, with similar pharmacophoric interaction fields, we consider this a success.

Dihydrofolate Reductase (DHFR). The target pharmacophore identified by Patel et al. consists of a donor, acceptor, and two hydrophobic regions. Figure 3A illustrates the FLAPpharm derived target pharmacophore. The hydrophobic interaction

dx.doi.org/10.1021/ci300153d | J. Chem. Inf. Model. 2012, 52, 2587−2598

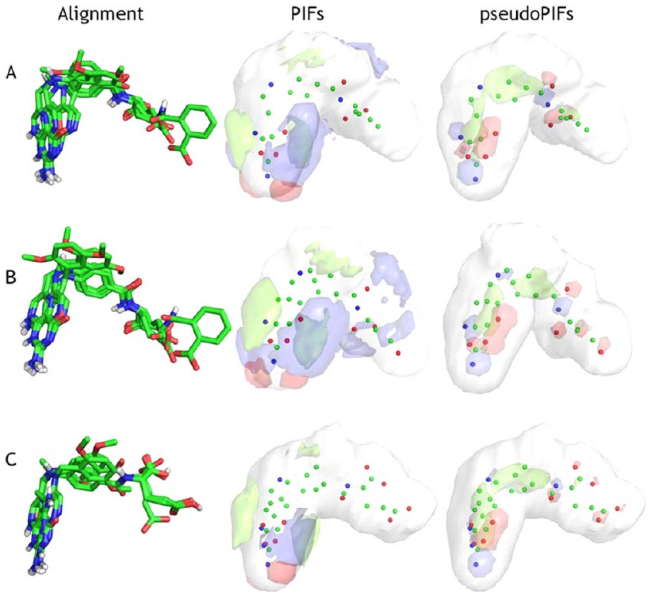


Figure 3. DHFR ligand alignments, and their respective pharmacophoric pseudomolecules. The pseudomolecules are shown twice, with either their pharmacophoric interaction fields (PIFs) or pharmacophoric pseudofields (pseudoPIFs) displayed, for clarity. The ligands are taken from (A) the X-ray-derived alignements, (B) the FLAPpharm rigid alignments, and (C) the FLAPpharm unbiased model alignments. Hydrophobic atoms and fields are displayed in green, hydrogen-bond donor atoms and fields are displayed in blue, hydrogen-bond acceptor atoms and fields are displayed in red, and the shape field is shown in white.

fields around the pteridine ring system are identified; however, for the aligned phenyl rings only one face is deemed accessible and able to interact with the protein, an aspect that would not be illustrated by a feature-only representation. There is a broad donor field interacting with the pteridine ring system in the central part of the figure, a partial donor field interacting with it on the left-hand side of the figure, and another partial donor field interacting with the amide and ether oxygen atoms adjacent to the phenyl rings. There are two lobes of an acceptor field interacting with the NH_2 group at the bottom of the figure, and another partial acceptor common to a few of the molecules on the left-hand side of the figure, partially obscured by the hydrophobic and donor features there.

For the rigid alignment (Figure 3B), all of the key interaction fields from the X-ray alignment are identified, with the addition of a partial donor field interacting with more crisply aligned carboxylate acceptors. In general, the features are more crisply aligned than in the X-ray alignment, and this increases the intensity of several regions of the PIFs.

For the unbiased model (exluding 1OHK due to the conformational issues described above, Figure 3C), the key features are all identified with the exception of the partial features that exist in the "tail" region of the molcules. The classical overlay of the folate and other analogues according to their fields (and not the scaffold) is observed. Given the similarity of both the alignments and the pharmacophoric interaction fields to those present in the X-ray and rigid alignments, and the fact that these molecules are more flexible, we consider it an outstanding success to find this as a top-ranked solution.

HIV-1 Reverse Transcriptase (HIVRT). The target pharmacophore identified by Patel et al. consists solely of a hydrophobic region common to all 10 ligands, and this makes aligning these molecules correctly a more difficult case. Figure 4A shows the target pharmacophore identified by FLAPpharm

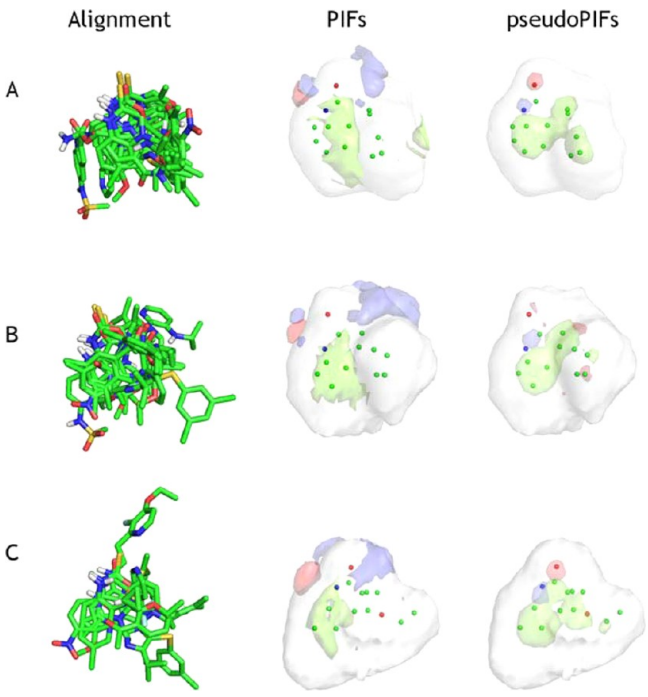


Figure 4. HIVRT ligand alignments and their respective pharmacophoric pseudomolecules. The pseudomolecules are shown twice, with either their pharmacophoric interaction fields (PIFs) or pharmacophoric pseudofields (pseudoPIFs) displayed, for clarity. The ligands are taken from (A) the X-ray-derived alignements, (B) the FLAPpharm rigid alignments, and (C) the FLAPpharm unbiased model alignments. Hydrophobic atoms and fields are displayed in green, hydrogen-bond donor atoms and fields are displayed in blue, hydrogen-bond acceptor atoms and fields are displayed in red, and the shape field is shown in white.

with these molecules. The hydrophobic region is identified and, additionally, a partial donor field that interacts with amide and thioamide acceptor atoms and the corresponding partial acceptor field that interacts with the amide and thioamide donor atoms, present in 6 of the 10 molcules.

For the rigid alignment (Figure 4B), all interactions from the X-ray alignments described above are found; however, several of the molecules are aligned incorrectly, in particular those four that do not share the common donor/acceptor motif described above. These incorrect alignments increase the magnitude of the hydrophobic PIF, and also some common acceptors are aligned to broaden the donor PIF at the top of the figure.

For the unbiased model (Figure 4C, with 1RT5 and 1KLM excluded due to the conformational reasons), the pharmacophore is almost identical to that found from the X-ray and rigid alignments. 1DTT is mis-aligned such that both of the thiourea donor atoms' corresponding acceptor fields overlay with the common acceptor field in the trans, trans configuration. An additional (incorrect) partial donor field is also found by aligning some of the noninteracting acceptor atoms. The pseudofields found corresepond directly to the interaction fields. In general, we consider this a fairly difficult case; the only feature common to all ligands is the hydrophobic region, with subsets of the molecules making alternative different inter-

dx.doi.org/10.1021/ci300153d | J. Chem. Inf. Model. 2012, 52, 2587−2598

actions with the protein. The main subset is those molecules that contain the amide/thioamide type donor/acceptor interaction, the minor subset is those that do not (1EP4, 1KLM, 1RT3, 1VRU). 1KLM was removed from the model building process; however, including the other three did not significantly affect the elucidation of the target pharmacophore. We consider this an advantage of the method. Another positive aspect is shown by the crisp atomic alignments of the amide/thioamide groups; a field-based alignment method might be expected to give more fuzzy alignments of the atoms giving rise to these fields. However, this is not necessarily the case when using the entire field, as opposed to using extension points, or even field sampling approaches.

Thermolysin. The target pharmacophore identified by Patel et al. consists of two acceptors, a hydrophobe, and a metal-binding feature which consists of two acceptors in the six ligands. Figure 5A shows FLAPpharm identified target

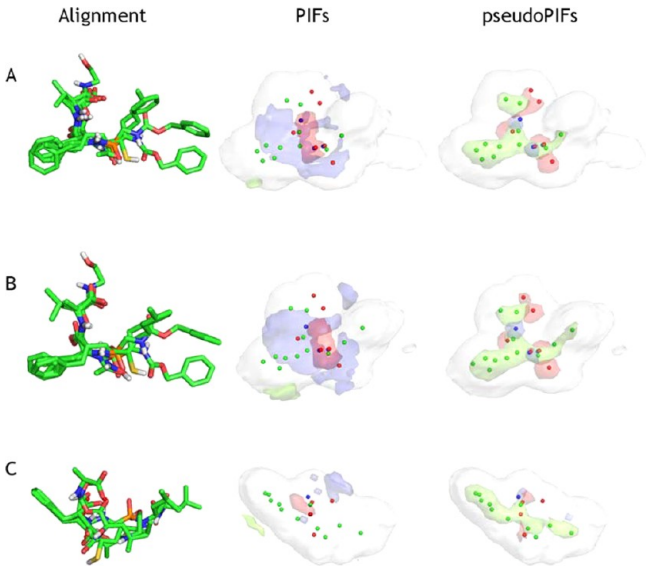


Figure 5. Thermolysin ligand alignments, and their respective pharmacophoric pseudomolecules. The pseudomolecules are shown twice, with either their pharmacophoric interaction fields (PIFs) or pharmacophoric pseudofields (pseudoPIFs) displayed, for clarity. The ligands are taken from (A) the X-ray-derived alignements, (B) the FLAPpharm rigid alignments, and (C) the FLAPpharm unbiased model alignments. Hydrophobic atoms and fields are displayed in green, hydrogen-bond donor atoms and fields are displayed in blue, hydrogen-bond acceptor atoms and fields are displayed in red, and the shape field is shown in white.

pharmacophore. The hydrophobe is identified (although only one face of the overlaid phenyl rings is deemed accessible), and the acceptors and metal binding region merge together in one large donor field. Additionally, an acceptor field corresponding to the overlaid amide and phosphonamide donor atoms is identified, which potentially makes a hydrogen-bonding interaction to the asparagine and alanine acceptors in the protein target.

Figure 5B shows the rigid alignment; all of the target features are identified, and in fact, it is almost impossible to recognize the difference between the ligands aligned according to the protein backbone atoms and those aligned according to the fields by FLAPpharm. The largest difference is that the FLAPpharm alignments are more "crisply" overlaid, as was shown in some of the HIVRT alignments discussed above.

For the unbiased alignment (Figure 5C, 4TMN removed due to the conformational issues described above), the method has not picked out the correct conformers and alignments; the ligands are in general more "twisted", 5TMN is flipped so that the phenyl group is making the hydrophobic interaction, and its amide acceptor oxygen is making the interaction where the metal is placed, instead of the phosphonate. The same features are found in the model pharmacophore, but the broad donor field is much reduced in intensity, and in general, the model is incorrect. The molecules in this data set are quite flexible, without a more constrained ligand to help focus the search space, and this is most likely the reason why the optimal solution is not found, especially since the rigid alignment described above was so good.

Thrombin. The target pharmacophore identified by Patel et al. consists of a basic group and three hydrophobic regions.

Figure 6A illustrates target pharmacophore identified by FLAPpharm. The three hydrophobic regions give rise to four

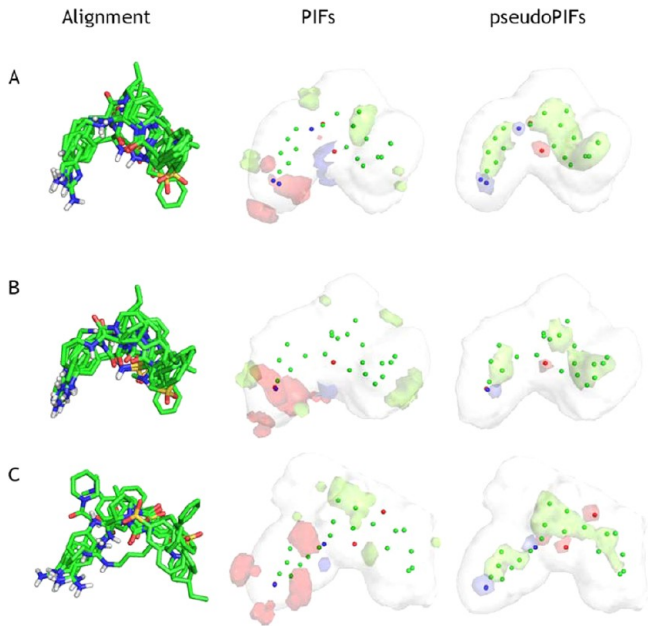


Figure 6. Thrombin ligand alignments and their respective pharmacophoric pseudomolecules. The pseudomolecules are shown twice, with either their pharmacophoric interaction fields (PIFs) or pharmacophoric pseudofields (pseudoPIFs) displayed, for clarity. The ligands are taken from (A) the X-ray-derived alignements, (B) the FLAPpharm rigid alignments, and (C) the FLAPpharm unbiased model alignments. Hydrophobic atoms and fields are displayed in green, hydrogen-bond donor atoms and fields are displayed in blue, hydrogen-bond acceptor atoms and fields are displayed in red, and the shape field is shown in white.

hydrophobic PIF regions, and the basic center is captured via the strong acceptor probe interactions. Additionally, partial interactions are shown interacting with an acceptor feature common to all but 1D4P (donor field).

Figure 6B shows the rigid alignment; the alignments and pharmacophoric interaction fields are close to those in the target pharmacophore, with the main difference being the accessible hydrophobic region, which is shifted slightly with respect to the X-ray alignments.

For the unbiased model (Figure 6C, with 1D6W removed due to the reasons identified earlier), the basic group is found as the intense acceptor field. According to the pharmacophoric

dx.doi.org/10.1021/ci300153d | J. Chem. Inf. Model. 2012, 52, 2587−2598

interaction fields, only one of the three hydrophobic regions are identified; however according to the pseudofields, all three are found. From the interaction field perspective, the hydrophobic probe interacts primarily with the two faces of each phenyl ring. The rings are aligned, but not cleanly; hence, the mean hydrophobic interaction energy is too low to be identified in the hydrophobic PIF. From the pseudoPIF perspective, the hydrophobic atoms are overlaid well; hence, the mean pseudofields are identified. The partial donor field present in the center of the figure is also identified, as in the target pharmacophore. There is also an additional acceptor pseudofield found toward the top-right-hand side of the figure, coming from three overlaid sp2 oxygen atoms. Therefore, we consider this a partial success.

FLAPpharm Pharmacophore Elucidation and Virtual Screening. As well as helping to understanding how ligands may bind to a common target, a key application of pharmacophore elucidation is to generate models that can then be used for virtual screening, i.e. to find similarly active molecules from a large pool of potentially available compounds. Previously we have reported the performance of FLAP in virtual screening,[22] using the DUD data set[24] that was designed to measure retrospective screening performance and enable comparison with other methods. In that study, given the flexibility of the FLAP approach, we tested many approaches, of which a brief description of a few of them is necessary to compare with the FLAPpharm pharmacophore screening approach. We used the chemotype clustering performed by Good[25] and focused on the data sets which contained 15 or more chemotypes to enable comparison with the work of Cheeseright et al.,[31] and we measured performance by analyzing what proportion of the chemotypes were retrieved during the search. The decoy set used was the "DUD-own" decoys, i.e. the more difficult set where the decoys were chosen to more closely represent the features present in the known actives. Using this approach, we felt it reasonable to test and compare ligand-based and receptor-based approaches (without chemotype clustering there is a bias to retrieving the many analogues present, especially for the ligand-based approaches). The first of the ligand-based approaches we used was simply to take the native cocrystallized ligand and use this as a template for FLAP, aligning and comparing all other molecules to this and scoring using the MIF similarity (LBX approach). The next approach is the receptor-based approach, aligning molecules into the receptor (akin to docking) and scoring how the molecules matched the MIFs of the target (RB approach). The third approach is similar to the RB approach, but in this case the receptor MIFs are filtered according to the bound ligand, i.e. only the key interacting fields are retained (RBLB approach). For FLAPpharm, we decided to use the 2D structures of the chemotype cluster parents to build models, and the top-ranked model was then used to align and score all of the other molecules in that data set. The results are shown in Figure 7, along with the equivalent performance from two other published methods.[24,30] The least complex FLAP ligand-based and receptor-based approaches (LBX and RB) perform almost equivalently, which is encouraging since receptor-based approaches are inherently more complex. Both of these "basic" FLAP approaches perform similarly to the ligand-based FieldScreen method, and all three outperform the original results given by DOCK. It is perhaps unsurprising that focusing the receptor MIFs to those that are more relevant from the perspective of the native ligand (RBLB approach)

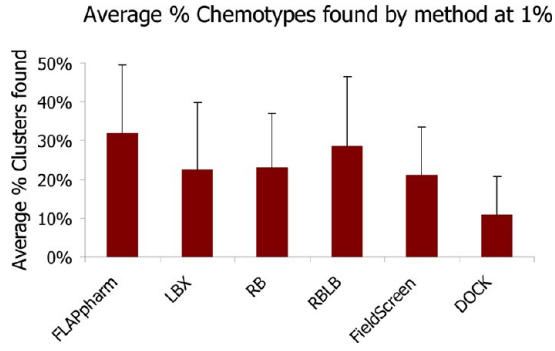


Figure 7. Virtual screening performance using the DUD data set, for several FLAP approaches, and compared to FieldScreen and DOCK.

increases performance, and in fact incorporating more information in the final molecule rankings through a variety of data fusion approaches increased the screening performance even more dramatically (although at the expense of performing several independent calculations).[22] However, for FLAPpharm, we believe the performance to be remarkable—in the absence of any 3D information—simply using the top-ranked pharmacophoric model without any inspection or refinement gave better results than the RBLB method which required a crystal structure of the target and ligand to be available. In the graphical interface implementation of FLAPpharm, it is also possible to remove certain features (for example if it is clear they are coming from the chemistry and not likely to be interacting with the target) and add specific constraints to force some regions to overlay. Using these software features, on a number of in-house targets, we have anecdotally observed even better performance (again reinforcing that additional knowledge can significantly improve results); however, we believe that the "baseline" we have presented here is already a huge step forward in this area.

Factor Xa Case Study. An intriguing case study to evaluate the FLAPpharm performance is represented by factor Xa (fXa). Inhibition of the blood serine protease fXa is of central importance for the treatment and prevention of thrombotic diseases. A number of X-ray structures are available for human fXa as a target, and we decided to use these to investigate the pharmacophore modeling approach in terms of characterizing the key features and virtual screening.

The canonical fashion binding mode for fXa inhibitors is characterized by two main interaction sites, named S1 and S4. The S1 site is bordered by hydrophobic walls and terminates with the negative charged aspartate.[32,33] The S4 pocket comprises mostly aromatic residues forming an "aromatic box".[34]

A typical fXa inhibitor usually fits into the binding site according to the following rules: (1) a positive charged group, usually a benzamidine group, interacts with the S1 site through hydrophobic interactions and forming a salt bridge with D189, and (2) an aromatic portion of the ligand interacts with the S4 pocket.

Molecules that contain surrogates of the benzamidine group do not make direct hydrogen-bonding interactions with D189 at the bottom of the S1 pocket. The main interactions involve hydrogen-bonding with the G219 carbonyl oxygen in the S1 subsite, a hydrophobic contact in the upper part of the S1 pocket, hydrogen-bonding with the G219 nitrogen amide, and aromatic contacts in the S4 cleft.[35]

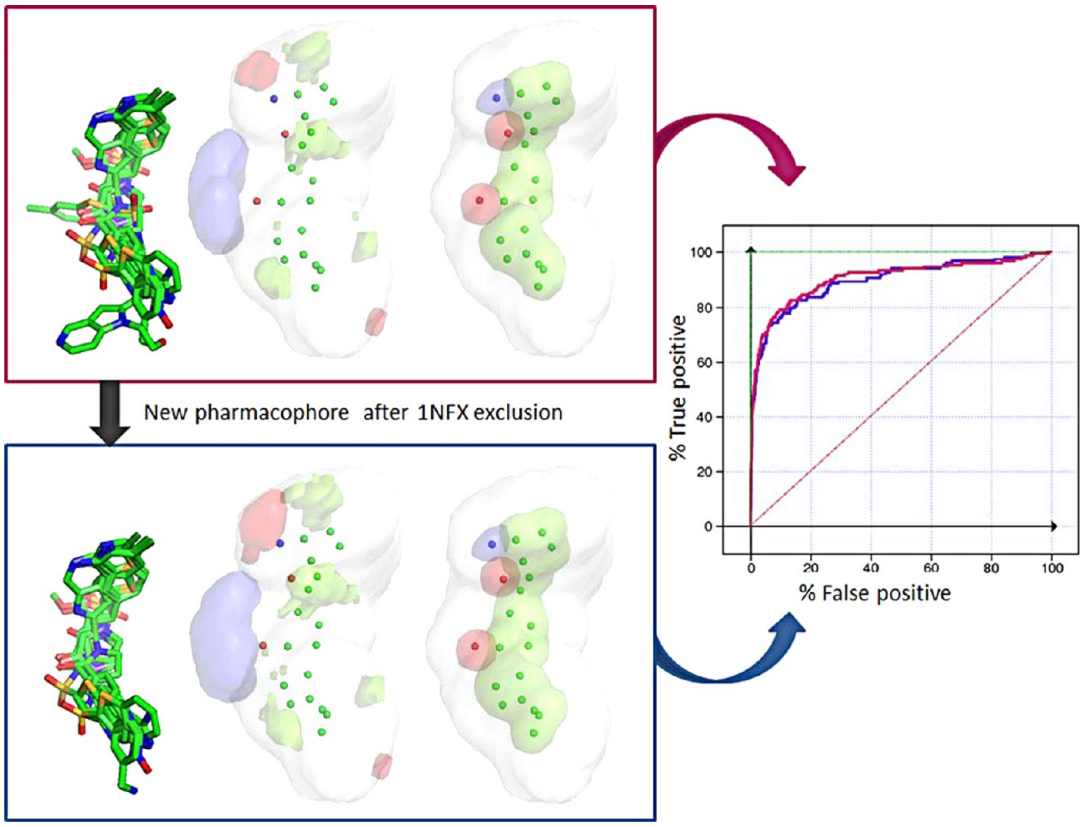


Figure 8. (left) fXa ligand alignments and their respective pharmacophoric pseudomolecules. The pseudomolecules are shown twice, with either their pharmacophoric interaction fields (PIFs) or pharmacophoric pseudofields (pseudoPIFs) displayed, for clarity. Hydrophobic atoms and fields are displayed in green, hydrogen-bond donor atoms and fields are displayed in blue, hydrogen-bond acceptor atoms and fields are displayed in red, and the shape field is shown in white. (right) ROC curve illustrating the virtual screening performance of the pharmacophore against the DUD data set.

Finally, a "chloro binding mode" has also been reported[36−38] and in fact using GRID to calculate the MIFs using organic chlorine as the probe uncovers this alternative favorable interaction which is extremely significant; a permanent charged group would likely be detrimental to oral bioavailability, whereas a neutral group (such as the dichlorophenyl) would not.[38] Indeed, it was found that in some cases, positively charged groups, such as guanidines or benzamidines, can be located in the S4 pocket, with a dichlorophenyl, chlorobenzothiophenyl, or chlorothiophenyl substituent binding in the S1 pocket.

In our study, eight X-ray structures of the human fXa were selected, having different binding modes: 1ezq, 1f0r, 1f0s, 1ksn, 1nfu, 1nfw, 1nfx, and 1nfy. Our aim was to verify whether or not the FLAPpharm model was able to handle these alternative binding modes and reproduce the correct X-ray alignment, investigate the virtual screening enrichment performance of this pharmacophore, and finally dock the pharmacophoric pseudomolecule as an alternative approach to validating the model. The X-ray conformations of the ligands were aligned using FLAPpharm, generating the pharmacophoric pseudomolecule shown in Figure 8.

The 1nfx ligand was not perfectly aligned; therefore, we decided to remove this ligand and regenerate the pharmacophore to test the sensitivity of the approach to a single misalignment. A comparison of the pharmacophores generated using the whole data set or removing the misaligned ligand shows that they are almost identical—the method therefore seems not to be so sensitive to a poor alignment in terms of visualizing the key pharmacophoric features. These two pharmacophores were also used as templates for virtual screening using the DUD data set for fXa containing 142 ligands and 5102 decoys.[24] The results were also highly comparable, again illustrating the insensitivity of the approach to a single misalignment and also that this pharmacophore yields excellent performance when used as a template for virtual screening to differentiate between known actives and decoys.

If no cocrystallized ligand data is available, pharmacophore models can still be built from two-dimensional (2D) structures with the aim of reproducing the experimental binding modes (as we have shown above using the Patel data sets). In some cases however, the pharmacophore model may not be the top-ranked one, and it is therefore important to validate each of the potential models. One way to do this is to perform virtual screening using known actives and some decoy structures; another could be to see how well the pharmacophore model fits with the receptor active site. Given that for FLAP the pharmacophoric pseudomolecule can be used in the same manner as any small molecule ligand (and also includes shape in addition to the pharmacophoric features), we decided to investigate whether docking the factor Xa pharmacophore into the receptor could provide useful information.

The generated pharmacophore described above was docked into the binding site using FLAP, and the 10 top-ranked poses were inspected. The first observation was that the docking was

dx.doi.org/10.1021/ci300153d | J. Chem. Inf. Model. 2012, 52, 2587−2598

successful, providing some evidence that the model is a good one.

The two top-ranked poses are reported in Figure 9. The second observation is also interesting; FLAP was able to reproduce the two main binding modes: in A the protonated group (blue point) is oriented toward the D189 and in B the protonated group (blue point) is oriented toward the S4 pocket, having the hydrophobic moiety pointing toward the W228 in the S1 pocket, according to the chloro-binding mode fashion. This chloro-binding mode, which is of great significance in the design of fXa inhibitors, was indirectly predicted using GRID MIFs.[38] While GRID itself is limited to inspection and manual interpretation, FLAP and FLAPpharm reproduce this prediction in an automatic fashion and enable the prospective searching for other inhibitors using the same technology.

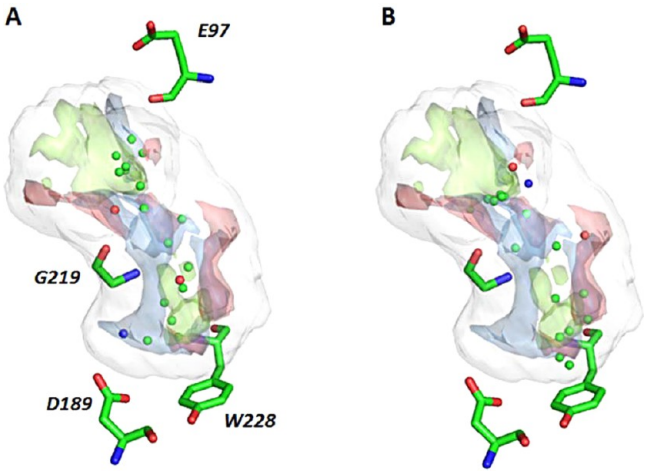


Figure 9. Factor Xa binding pocket with molecular interaction fields highlighting the key interactions with the receptor. (A and B) Top-ranked binding modes when the pharmacophoric pseudomolecule is docked into the receptor (pharmacophoric points: hydrophobic = green, acceptor = red, donor = blue; PIFs and pseudoPIFs not shown for clarity). See text for details.

CONCLUSIONS

Pharmacophore elucidation is mainly applied to (a) molecular alignment and SAR analysis, to understand how active molecules might bind to a common receptor site and (b) for the virtual screening of potentially active molecules from a large pool of either available or virtual compounds. Classical feature-based approaches suffer from several limitations: a template molecule may be required, the features can be simplistic, sterically inaccessible features may still be considered, the features may align well but the molecules themselves poorly, the number of allowed "misses" usually needs to be specified in advance by the user, and the number of partially matching features as well. A further drawback is that typically a large number of potential models is produced (in some cases 100s of models) which then need to be analyzed, refined, and validated by the user, which can be an expert task that is time-consuming. Here we have presented a new approach called FLAPpharm, that enables the elucidation of several pharmacophoric pseudomolecule hypotheses. To us, the common pharmacophore is better described as a pharmacophoric pseudomolecule consisting of pharmacophoric interaction fields (PIFs), atom-centered pharmacophoric pseudoMIFs (pseudoPIFs), and pharmacophoric points representing the centroid of these pseudoMIFs. The method elucidates the pharmacophoric pseudomolecules after hypothesizing the bioactive conformations and the subsequent common molecular alignment and suffers from none of the limitations outlined above for the feature-based approaches. We have shown that the conformational searching is capable of reproducing close to the bioactive conformation in greater than 90% of cases, that the method is able to align molecules to reproduce the X-ray derived alignments with great success, and that the derived alignments when starting from only 2D structural information are excellent, with only one of the five cases (thermolysin) producing an incorrect model due to the flexibility of the data set. The PIFs and pseudoPIFs capture the key interacting features from the target perspective as well as the chemist's perspective and inherently include partial matching. The top-ranked model is the best in all of the cases here, although anecdotally we have seen other cases where it is prudent to examine the top five alignment models. Calculation time is of course increased compared to using the rule-based feature approaches; however, we believe it is still reasonable: single conformer alignments are usually performed in under a minute, a full search considering up to 30 conformations can take several minutes or an hour depending on the number of molecules (100 input molecules is the largest data set we have modeled). It is our opinion that the required time to analyze the models is however reduced significantly (for example in the virtual screening study no analysis was performed at all and the top-ranked models used in an automated approach). To FLAP, the pharmacophoric pseudomolecules can be used as templates for virtual screening, or they can be docked into a target site to help validate the hypothesis; the alignments may also be used for 3D-QSAR. We have also demonstrated here that the models are extremely useful for virtual screening; starting only from 2D information about a subset of active molecules, the top-ranked FLAPpharm hypotheses outperformed all of the other approaches. In the final case study, we have shown how a FLAPpharm hypothesis might be constructed for a set of fXa inhibitors and that the top-ranked model is insensitive to a single misaligned structure. Subsequent validation of the model using virtual screening and also docking into the target receptor illustrate two alternative approaches to postprocessing and selecting between potential alternative pharmacophore hypotheses. From the docking, the resulting two top-ranked poses illustrated the most common experimentally known binding modes, including prediction of the important alternative "chloro-binding" mode which is of great significance in the design of fXa inhibitors with improved oral bioavailability. As such we believe FLAPpharm shows great promise in the area of pharmacophore modeling and again illustrates the utility of GRID molecular interaction fields in the area of in silico drug discovery. Part II of this paper presents a comprehensive data set of experimentally derived molecular alignments for 81 different target sets, with the aim of providing a benchmark data set for pharmacophore elucidation in terms of molecular alignment quality and analogous to the DUD data set described here for virtual screening performance.[26]

AUTHOR INFORMATION

Corresponding Author

*E-mail: simon@moldiscovery.com.

Notes

The authors declare no competing financial interest.

REFERENCES

(1) Van Drie, J. H. Monty Kier and the origin of the pharmacophore concept. *Internet Electron. J. Mol. Des.* **2007**, *6*, 271−279.

(2) Marshall, G. R.; Barry, C. D.; Bosshard, H. E.; Dammkoehler, R. A.; Dunn, D. A. The Conformational Parameter in Drug Design: The Active Analog Approach. In *Computer-Assisted Drug Design*; Olson, E. C., Christoffersen, R. E., Eds; ACS Symposium Series 112; American Chemical Society: Washington, DC, 1979; pp 205−226.

(3) Leach, A. R.; Gillett, V. J.; Lewis, R. A.; Taylor, R. Three-Dimensional Pharmacophore Methods in Drug Discovery. *J. Med. Chem.* **2010**, *53*, 539−558.

(4) Langer, T.; Wolber, G. Pharmacophore definition and 3D searches. *Drug Discovery Today: Technologies* **2004**, *1*, 203−207.

(5) Goodford, P. J. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J. Med. Chem.* **1985**, *28*, 849−857.

(6) Cruciani, G.; Pastor, M.; Guba, W. VolSurf: a new tool for the pharmacokinetic optimization of lead compounds. *Eur. J. Pharm. Sci.* **2000**, *11*, S29−S39.

(7) Cruciani, G.; Carosati, E.; De Boeck, B.; Ethirajulu, K.; Mackie, C.; Howe, T.; Vianello, R. MetaSite: Understanding Metabolism in Human Cytochromes from the Perspective of the Chemist. *J. Med. Chem.* **2005**, *48*, 6970−6979.

(8) Milletti, F.; Storchi, L.; Sforna, G.; Cruciani, G. New and Original pK_a Prediction Method Using Grid Molecular Interaction Fields. *J. Chem. Inf. Model.* **2007**, *47*, 2172−2181.

(9) Milletti, F.; Storchi, L.; Sforna, G.; Cross, S.; Cruciani, G. Tautomer enumeration and stability prediction for virtual screening on large chemical databases. *J. Chem. Inf. Model.* **2009**, *49*, 68−75.

(10) Bergmann, R.; Linusson, A.; Zamora, I. SHOP: Scaffold HOPping by GRID-Based Similarity Searches. *J. Med. Chem.* **2007**, *50*, 2708−2717.

(11) Fontaine, F.; Pastor, M.; Zamora, I.; Sanz, F. Anchor-GRIND: Filling the gap between standard 3D QSAR and the GRid-Independent descriptors. *J. Med. Chem.* **2005**, *48*, 2687−2694.

(12) Crivori, P.; Cruciani, G.; Carrupt, P. A.; Testa, B. Predicting blood-brain barrier permeation from three-dimensional molecular structure. *J. Med. Chem.* **2000**, *43*, 2204−2216.

(13) Ahlström, M. M.; Ridderström, M.; Zamora, I.; Luthman, K. CYP2C9 Structure-Metabolism Relationships: Optimizing the Metabolic Stability of COX-2 Inhibitors. *J. Med. Chem.* **2007**, *50*, 4444−4452.

(14) Carosati, E.; Mannhold, R.; Wahl, P.; Hansen, J. B.; Fremming, T.; Zamora, I.; Cianchetta, G.; Baroni, M. Virtual screening for novel openers of pancreatic KATP channels. *J. Med. Chem.* **2007**, *50*, 2117−2126.

(15) Broccatelli, F.; Carosati, E.; Neri, A.; Frosini, M.; Goracci, L.; Oprea, T. I.; Cruciani, G. A Novel Approach for Predicting P-glycoprotein (ABCB1) Inhibition Using Molecular Interaction Fields. *J. Med. Chem.* **2011**, *54*, 1740−1751.

(16) Cross, S; Cruciani, G. Molecular fields in drug discovery: getting old or reaching maturity? *Drug Discov. Today* **2010**, *15*, 23−32.

(17) Cross, S; Cruciani, G. Grid-derived structure-based 3D pharmacophores and their performance compared to docking. *Drug Discov. Today: Technol.* **2010**, *7*, e213−e219.

(18) Baroni, M.; Cruciani, G.; Sciabola, S.; Perruccio, F.; Mason, J. S. A Common Reference Framework for Analyzing/Comparing Proteins and Ligands. Fingerprints for Ligands And Proteins (FLAP): Theory and Application. *J. Chem. Inf. Model.* **2007**, *47*, 279−294.

(19) Carosati, E.; Cruciani, G.; Chiarini, A.; Budriesi, R.; Ioan, P.; Spisani, R.; Spinelli, D.; Cosimelli, B.; Fusi, F.; Frosini, M.; Matucci, R.; Gasparrini, F.; Ciogli, A.; Stephens, P. J.; Devlin, F. J. Calcium Channel Antagonists Discovered by a Multidisciplinary Approach. *J. Med. Chem.* **2006**, *49*, 5206−5216.

(20) Carosati, E.; Sforna, G.; Pippi, M.; Marverti, G.; Ligabue, A.; Guerrieri, D.; Piras, S.; Guaitoli, G.; Luciani, R.; Costi, M. P.; Cruciani, G. Ligand-Based Virtual Screening and ADME-Tox Guided Approach to Identify Triazolo-quinoxalines as Folate Cycle Inhibitors. *Bioorg. Med. Chem.* **2010**, *18*, 7773−7785.

(21) Brincat, J. P.; Carosati, E.; Sabatini, S.; Manfroni, G.; Fravolini, A.; Raygada, J. L.; Patel, D.; Kaatz, G. W.; Cruciani, G. Discovery of Novel Inhibitors of the NorA Multidrug Transporter of Staphylococcus aureus. *J. Med. Chem.* **2011**, *54*, 354−365.

(22) Cross, S.; Baroni, M.; Carosati, E.; Benedetti, P.; Clementi, S. FLAP: GRID Molecular Interaction Fields in Virtual Screening. Validation using the DUD Data Set. *J. Chem. Inf. Model.* **2010**, *50*, 1442−1450.

(23) Patel, Y.; Gillet, V. J.; Bravi, G.; Leach, A. R. A comparison of the pharmacophore identification programs: Catalyst, DISCO and GASP. *J. Comput.-Aided Mol. Des* **2002**, *16*, 653−681.

(24) Huang, N.; Shoichet, B. K.; Irwin, J. J. Benchmarking Sets for Molecular Docking. *J. Med. Chem.* **2006**, *49*, 6789−6801.

(25) Good, A. C.; Oprea, T. I. Optimization of CAMD techniques 3. Virtual screening enrichment studies: a help or hindrance in tool selection? *J. Comput.-Aided Mol. Des.* **2008**, *22*, 169−178.

(26) Cross, S.; Ortuso, F.; Baroni, M.; Costa, G.; Distinto, S.; Moraca, F.; Alcaro, S.; Cruciani, G. GRID-Based Three-Dimensional Pharmacophores II: PharmBench, a Benchmark Dataset for Evaluating Pharmacophore Elucidation Methods. *J. Chem. Inf. Model.* **2012**, DOI: 10.1021/ci300154n.

(27) Allinger, N. L.; Yuh, Y. H.; Lii, J.-H. Molecular mechanics. The MM3 force field for hydrocarbons. 1. *J. Am. Chem. Soc.* **1989**, *111*, 8551−8566.

(28) Allen, F. H. The Cambridge Structural Database: a quarter of a million crystal structures and rising. *Acta Crystallogr.* **2002**, *B58*, 380−388.

(29) Wang, R.; Fang, X.; Lu, Y.; Wang, S. The PDBbind Database: Collection of Binding Affinities for Protein−Ligand Complexes with Known Three-Dimensional Structures. *J. Med. Chem.* **2004**, *47*, 2977−2980.

(30) Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Delivery Rev.* **1997**, *23*, 3−25.

(31) Cheeseright, T. J.; Mackey, M. D.; Melville, J. L.; Vinter, J. G. FieldScreen: Virtual Screening Using Molecular Fields. Application to the DUD Data Set. *J. Chem. Inf. Model.* **2008**, *48*, 2108−2117.

(32) Padmanabhan, K.; Padmanabhan, K. P.; Tulinsky, A.; Park, C. H.; Bode, W.; Huber, R.; Blankenship, D. T.; Cardin, A. D.; Kisiel, W. Structure of Human Des(1−45) Factor Xa at 2·2 Å Resolution. *J. Mol. Biol.* **1993**, *232*, 947−966.

(33) Brandstetter, H.; Kuhne, A.; Bode, W.; Huber, R.; Von der Saal, W.; Wirthensohn, K.; Engh, R. X-ray Structure of Active Site-inhibited Clotting Factor Xa: Implications for drug design and substrate recognition. *J. Biol. Chem.* **1996**, *271*, 29988−29992.

(34) Bhunia, S. S.; Roy, K. K.; Saxena, A. K. Profiling the structural determinants for the selectivity of representative factor-Xa and thrombin inhibitors using combined ligand-based and structure-based approaches. *J. Chem. Inf. Model.* **2011**, *51*, 1966−1985.

(35) Maignan, S.; Guilloteau, J.-P.; Pouzieux, S.; Choi-Sledeski, Y. M.; Becker, M. R.; Klein, S. I.; Ewing, W. R.; Pauls, H. W.; Spada, A. P.; Mikol, V. Crystal Structures of Human Factor Xa Complexed with Potent Inhibitors. *J. Med. Chem.* **2000**, *43*, 3226−3232.

(36) Maignan, S.; Guilloteau, J.-P.; Choi-Sledeski, Y. M.; Becker, M. R.; Ewing, W. R.; Pauls, H. W.; Spada, A. P.; Mikol, V. Molecular structures of human factor Xa complexed with ketopiperazine inhibitors: Preference for a neutral group in the S1 pocket. *J. Med. Chem.* **2003**, *46*, 685−690.

(37) Adler, M.; Kochanny, M. J.; Ye, B.; Rumnenik, G.; Light, D. R.; Biancalana, S. Whitlow, M. Crystal Structures of two potent nonamidine inhibitors bound to factor Xa. *Biochemistry* **2002**, *41*, 15514−15523.

(38) Matter, H.; Will, D. W.; Nazaré, M.; Schreuder, H.; Laux, V.; Wehner, V. Structural Requirements for Factor Xa Inhibition by 3-Oxybenzamides with Neutral P1 Substituents: Combining X-ray

Crystallography, 3D-QSAR, and Tailored Scoring Functions. *J. Med. Chem.* **2005**, *48*, 3290−3312.