JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

Assessment of a Cambridge Structural Database-Driven Overlay Program

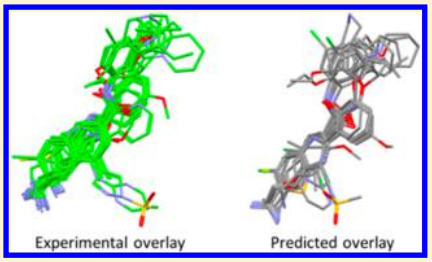
Ilenia Giangreco,*[,†] Tjelvar S. G. Olsson,[†] Jason C. Cole,[†] and Martin J. Packer[‡]

[†]Cambridge Crystallographic Data Centre, 12 Union Road, Cambridge CB2 1EZ, United Kingdom

[‡]AstraZeneca, Mereside, Alderley Park, Macclesfield SK10 4TG, United Kingdom

Supporting Information

ABSTRACT: We recently published an improved methodology for overlaying multiple flexible ligands and an extensive data set for validating pharmacophore programs. Here, we combine these two developments and present evidence of the effectiveness of the new overlay methodology at predicting correct superimpositions for systems with varying levels of complexity. The overlay program was able to generate correct predictions for 95%, 73%, and 39% of systems classified as easy, moderate, and hard, respectively.



INTRODUCTION

Pharmacophore elucidation aims to identify the set of key common features and their relative position (the pharmacophore model) from a set of molecules that are known to bind to a given therapeutic target. The pharmacophore model can be used to rationalize and, if correct, predict the effect of a chemical structure modification to a given binder, as the model gives the researcher an understanding for the requirement of chemical features in 3D space. In practice, a given set of ligands may each bind in a different way to the same target (either in multiple different binding modes to the same pocket or via allosteric binding) or may bind via solvent-mediated interactions that may invert the donor−acceptor behavior of a given functional group, such situations present challenges for overlay programs, as typically the ligands are considered in isolation.

In 2010, Leach et al.[1] reviewed the use of 3D pharmacophore methods in drug discovery raising a number of issues and highlighting some areas for improvements. They acknowledged the difficulty in aligning molecules, not just because of any inherent limitations of the programs available but because the problem is underdetermined; typically, in the absence of additional information about binding, there may be multiple overlays that present legitimate hypotheses for binding. In order to address this problem, our program was designed to produce a diverse set of plausible molecular overlays that all give potential pharmacophore hypotheses to test. The overlays are scored with three objective functions measuring the overlay quality (union volume, hydrogen-bond match, and hydrophobic match), filtered on Pareto rank,[2] and sorted in ascending order based on their Borda tallies for the three objective functions.[3]

Another aspect pointed out by Leach et al.[1] was the need to improve validation standards by creating good data sets of "true" molecular overlays with which to evaluate new pharmacophore programs or perform comparative studies of different programs. To this end, a group in AstraZeneca (AZ) recently published an extensive and diverse benchmarking set from high quality protein−ligand complexes.[4] This test set contains 121 experimentally derived molecular overlays spanning across multiple protein families. This set can be regarded as a benchmarking standard to demonstrate the ability of a given program to reproduce experimental data.

In this paper, we present an extended scientific assessment of a CSD-driven overlay program using this data set. The original publication describing the overlay methodology already included a validation study using 10 ligand sets.[5]

MATERIALS AND METHODS

Benchmarking Data Set. The overlay program was assessed by comparing its predictions with 121 molecular overlays in what we will call the AZ test set.[4] Full details about the rules applied to select protein−ligand complexes, the selection of residues used to align the protein structures, and the selection of an appropriately sized subset are described in the original paper.

Starting from the assumption that a good pharmacophore model should explain how structurally diverse ligands can bind to a common receptor site, the AZ test set was annotated with three parameters indicating the overlay quality: (1) The average shape match, calculated on all the possible pairwise combinations of ligands within a set.[6] (2) The average Color score, which accounts for feature similarity. The Color score is a measure of overlap for predefined pharmacophore points obtained by aligning groups with the related properties, as contained in the Color force field definition file.[7] (3) The average Tanimoto coefficient used to measure 2D fingerprint similarity.[8]

Received: August 20, 2014
Published: November 13, 2014

A consensus of these parameters was used to rank the data sets based on their suitability for pharmacophore elucidation assessment.

In this work, we evaluated the performance of the program against all systems in the AZ test set. We therefore reranked the AZ test set based on the level of difficulty in predicting the observed ("true") overlay of a given set of ligands, rather than on how suitable it was for validation. In other words, a set where all ligands were similar to each other would be considered easy to predict rather than unsuitable for validation purposes.

The new ranking approach used the same parameters as the suitability score, but sorting of the average Tanimoto coefficient was reversed. We calculated the Borda tallies for the three parameters, that is, shape matching, Color score, and Tanimoto coefficient, and ranked them in ascending order. The lower the Borda tally, the easier the prediction should be for this data set. A distribution plot of Borda tallies, shown in Figure 1, was used

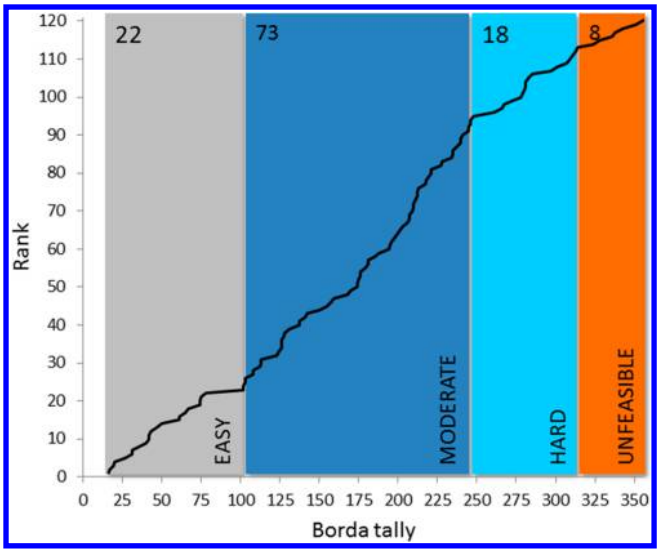


Figure 1. Distribution of the Borda tallies of the three parameters used to rank the AZ test set based on the difficulty to reproduce the experimental overlay. The colors indicate our criterion of classification. Changes in gradient were used to suggest the boundaries between the various categories. This was confirmed by subsequent visual inspection.

as a guideline for the classification, and the boundaries between the various categories were defined by the changes in gradient. A data set was classified as easy if the Borda tally was lower than or equal to 100, moderate if the Borda tally was between 100 and 250, hard if the Borda tally was between 250 and 314, and unfeasible if the Borda tally was greater than or equal to 314. Eight systems were deemed unfeasible due to their poor Color scores; visual inspection confirmed that information contained in these systems was insufficient to align them. This assignment agreed with the original ranking of the AZ test set based on the suitability of each set for validation study, where these systems scored at the lower end of the range.

Table 1 shows the AZ test set ranked and classified with this new approach.

Conformer Library Generation. Ligand models were created with CORINA[9] and used as input for conformer generation in order to avoid possible biases that could occur by starting with the X-ray conformation. An in-house tool (a CSD-driven conformer generator, to be published separately) was used to generate conformer libraries for all ligands. This approach makes use of ring templates and distributions of bond lengths, angles, and rotamer geometries derived from the Cambridge Structural Database (CSD)[10] and retrieved from Mogul.[11] In the first step, the molecule's bond lengths and angles are optimized to give the best fit to the CSD distributions using a gradient-based minimization technique in Cartesian space. The molecule is then disassembled into rigid components connected by rotatable bonds. For each rotatable bond, a matching rotamer distribution is used to determine likely torsion angle settings, and for components containing flexible ring systems, matching ring templates in the CSD were identified.[12] Each setting, that is, a rotamer torsion angle or a ring template for a flexible ring, is weighted by the number of observations of similar geometries in the CSD. Given this information, a tree sampling approach of up to one million conformers is constructed in memory, which is then sorted according to overall likelihood. The sorted list is finally clustered to select up to 200 conformations using a geometric dissimilarity approach.

CSD-Driven Overlay Program. The algorithm returns multiple diverse overlays as plausible solutions to be used in pharmacophore elucidation studies and virtual screening campaigns. The following sequence describes the key steps of the methodology:

(1) *Overlay generation*: A fingerprint technique that generates several thousand possible overlays using bit-string manipulations.

(2) *Overlay filtering*: Each overlay is scored on three objective functions (union volume, hydrogen-bond match, and hydrophobic match), and ranked by constrained Pareto ranking. A diverse subset of the best solutions is then chosen using an overlay dissimilarity metric. After filtering, 20 overlays are retained.

Because the overlay algorithm uses stochastic searches, a panel of calculations was set up to evaluate the repeatability of the results. We ran 10 jobs for each set of ligands using default settings.

Performance Evaluation and Success Criteria. Establishing exact criteria to define success is one of the main problems when validating pharmacophore programs. Recently, FLAPpharm[13] was assessed using *AlignScore* as a metric; this metric was first introduced by Jones to validate GAPE.[14] The *AlignScore* criterion for success is that at least half of the ligands in the predicted overlay can be fitted to their equivalents in the experimental overlay with an average heavy atom root-mean-square deviation (RMSD) $\leq$ 2 Å.

In this work, we changed the termination criterion to calculate *AlignScore*, making it more stringent. The modified algorithm makes use of RMSD values for the individually fitted molecules, which needs to be $\leq$ 2 Å, rather than the average RMSD of the ligands in the set being $\leq$ 2 Å (full details are provided in the Supporting Information). An implicit constraint of this modified algorithm is that any overlay that passes the new termination criterion will also pass the original one.

Selection of One Nondominated Overlay Solution. The overlay program returns multiple overlay solutions for each job, ranked in ascending order based on their Borda tallies for the three objective functions, which in this context are union volume, hydrogen-bond match, and hydrophobic match. Another key descriptor is the dominance value of the Pareto ranking; one solution dominates another if, and only if, it

Table 1. Classification of AZ Test Set[4] Based on the Level of Difficulty to Reproduce the Experimental Overlay*

UniProt ID	Target	Category	Is congeneric?	Avg success[a]	Avg success[b]	Avg success[c]
P05326	isopenicillin n synthase	easy	yes	10	10	7
P00797	renin	easy	yes	10	5	4
P0ABP9	purine nucleoside phosphorylase	easy	yes	10	10	10
P30291	wee1-like protein kinase	easy	yes	10	10	10
P00489	protein (glycogen phosphorylase)	easy	no	10	10	10
P51955	serine/threonine-protein kinase NEK2	easy	no	10	10	10
P54760	ephrin type-B receptor 4	easy	yes	0	0	0
P12758	uridine phosphorylase	easy	yes	9	1	5
Q57834	tyrosyl-tRNA synthetase	easy	yes	10	10	10
P0A017	dihydrofolate reductase	easy	yes	10	8	8
P23470	receptor-type tyrosine-protein phosphatase gamma	easy	yes	10	9	10
P07688	cathepsin B	easy	yes	10	8	5
Q10714	angiotensin converting enzyme	easy	yes	10	8	10
Q9BJF5	calmodulin-domain protein kinase 1	easy	yes	10	10	10
P22906	dihydrofolate reductase	easy	yes	10	10	10
P08235	mineralocorticoid receptor	easy	no	10	10	10
P00374	dihydrofolate reductase	easy	no	7	0	1
P56658	adenosine deaminase	easy	yes	10	9	6
P61823	pancreatic ribonuclease A	easy	yes	10	10	9
P36897	TGF-beta receptor type I	easy	yes	10	10	10
P16184	dihydrofolate reductase	easy	no	9	2	1
P00509	aspartate aminotransferase	easy	yes	10	7	9
P14324	farnesyl pyrophosphate synthetase	moderate	no	10	10	10
P00730	carboxypeptidase A	moderate	no	10	10	10
P00760	trypsin	moderate	no	10	10	10
Q04771	activin receptor type-1	moderate	no	10	3	0
Q00511	uricase	moderate	no	10	10	10
P00734	alpha thrombin	moderate	no	3	1	0
Q9T0N8	cytokinin dehydrogenase 1	moderate	no	10	10	10
Q581W1	pteridine reductase 1	moderate	no	10	10	10
P02829	HSP82	moderate	no	10	5	9
P09955	procarboxypeptidase B	moderate	no	2	0	0
Q02127	dihydroorotate dehydrogenase, mitochondrial	moderate	no	10	9	9
P08254	stromelysin-1	moderate	no	0	0	0
O60885	human BRD4	moderate	no	9	0	4
P00742	coagulation factor XA	moderate	no	0	0	0
P39900	macrophage metalloelastase	moderate	no	10	6	6
Q92731	estrogen receptor beta	moderate	no	10	4	4
P00929	tryptophan synthase	moderate	no	10	9	8
P03372	oestrogen receptor	moderate	no	10	6	2
O15530	3-phosphoinositide dependent protein kinase-1	moderate	no	5	2	0
P09467	fructose-1,6-bisphosphatase 1	moderate	yes	10	10	10
P04035	protein (HMG-COA reductase)	moderate	no	3	1	0
P00469	thymidylate synthase	moderate	yes	10	5	2
P0C5C1	beta-lactamase	moderate	no	8	0	1
P10275	androgen receptor	moderate	no	10	5	7
P51857	3-oxo-5-beta-steroid 4-dehydrogenase	moderate	no	10	10	10
O14965	serine/threonine-protein kinase 6	moderate	no	3	0	1
P25774	cathepsin S	moderate	no	0	0	0
P30405	peptidyl-prolyl cis-trans isomerase F, mitochondrial	moderate	no	10	10	10
P00918	carbonic anhydrase II	moderate	no	10	1	0
P24941	cyclin-dependent kinase 2	moderate	no	10	6	5
P15090	fatty acid-binding protein, adipocyte	moderate	no	10	10	10
P41148	endoplasmin	moderate	no	9	0	0
P08069	insulin-like growth factor 1 receptor precursor	moderate	yes	9	1	0
P35557	glucokinase isoform 2	moderate	no	10	10	10
P09960	leukotriene A-4 hydrolase	moderate	no	2	0	0
P04642	L-lactate dehydrogenase A chain	moderate	yes	10	10	10
P45452	collagenase 3	moderate	no	9	3	2
P56817	beta-secretase 1	moderate	no	0	0	0
P35968	vascular endothelial growth factor receptor 2	moderate	no	0	0	0
P07900	HSP 90-alpha	moderate	no	9	4	8
P08581	hepatocyte growth factor receptor	moderate	no	0	0	0
Q16539	p38 MAP kinase	moderate	no	0	0	0
P00749	protein (urokinase-type plasminogen activator)	moderate	no	10	8	10
P11509	cytochrome P450, family 2, subfamily A, polypeptide 6	moderate	no	10	10	10
Q9QYJ6	phosphodiesterase-10A	moderate	no	0	0	0
P43235	cathepsin K	moderate	no	6	2	0
P00517	cAMP-dependent protein kinase, alpha-catalytic subunit	moderate	no	0	0	0
P08709	coagulation factor VII	moderate	no	0	0	0
P25779	cruzain	moderate	no	0	0	0
Q07343	cAMP-specific 3',5'-cyclic phosphodiesterase 4B	moderate	no	1	0	0
O14757	serine/threonine-protein kinase Chk1	moderate	no	1	0	0
P0AE18	methionine aminopeptidase	moderate	no	8	3	1
P78536	ADAM 17	moderate	no	0	0	0

dx.doi.org/10.1021/ci500509y | J. Chem. Inf. Model. 2014, 54, 3091−3098

Table 1. continued

UniProt ID	Target	Category	Is congeneric?	Avg success[a]	Avg success[b]	Avg success[c]
Q13526	peptidyl-prolyl cis-trans isomerase NIMA-Interacting 1	moderate	no	0	0	0
Q9L5C8	beta-lactamase CTX-M-9	moderate	no	4	1	0
P0A5J2	methionine aminopeptidase	moderate	yes	4	0	4
P24182	biotin carboxylase	moderate	no	10	10	8
P25440	bromodomain-containing protein 2	moderate	no	10	6	8
P00520	proto-oncogene tyrosine-protein kinase ABL	moderate	no	4	0	0
O60674	tyrosine-protein kinase JAK2	moderate	no	0	0	0
P06401	progesterone receptor	moderate	no	0	0	0
O76290	pteridine reductase	moderate	no	10	8	6
P15121	aldose reductase	moderate	no	2	0	1
P00523	proto-oncogene tyrosine-protein kinase Src	moderate	no	2	0	0
Q9Y233	cAMP and cAMP-inhibited cGMP 3', 5'-cyclic phosphodiesterase 10A	moderate	no	10	4	2
P0AD64	beta-lactamase SHV-1	moderate	no	0	0	0
O76074	cGMP-specific 3',5'-cyclic phosphodiesterase	moderate	no	0	0	0
P27487	dipeptidyl peptidase IV soluble form	moderate	no	0	0	0
P18031	protein (protein-tyrosine phosphatase 1b)	moderate	no	4	0	0
P47811	mitogen-activated protein kinase 14	moderate	no	7	0	0
P17612	cAMP-dependent protein kinase	moderate	no	0	0	0
P28845	corticosteroid 11-beta-dehydrogenase isozyme 1	moderate	no	0	0	0
P28482	mitogen-activated protein kinase 1	moderate	no	10	4	7
Q08499	cAMP-specific 3',5'-cyclic phosphodiesterase 4D	hard	no	0	0	0
P53779	mitogen-activated protein kinase 10	hard	no	0	0	0
P06239	LCK kinase	hard	no	2	0	0
P28523	casein kinase II	hard	no	0	0	0
P49841	glycogen synthase kinase-3 beta	hard	no	0	0	0
P00772	elastase	hard	no	0	0	0
O15530 allosteric	3-phosphoinositide dependent protein kinase-2	hard	no	2	0	0
P80457	xanthine dehydrogenase	hard	no	0	0	0
P48736	phosphatidylinositol-4,5-bisphosphate 3-kinase	hard	no	9	8	1
P14324 allosteric	farnesyl pyrophosphate synthetase	hard	no	10	10	10
P11309	proto-oncogene serine/threonine-protein kinase Pim-1	hard	no	0	0	0
Q9BZP6	acidic mammalian chitinase	hard	no	0	0	0
P68400	casein kinase II	hard	no	10	6	4
P00808	beta-lactamase	hard	no	0	0	0
P52700	metallo-beta-lactamase L1	hard	no	10	4	2
P50579	protein (methionine aminopeptidase)	hard	no	0	0	0
P42330	aldo-keto reductase family 1 member C3	hard	no	0	0	0
Q3JRA0	2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase	hard	no	10	10	6
P04058	acetylcholinesterase	unfeasible	no	0	0	0
P42574	caspase-3	unfeasible	no	0	0	0
P14174	macrophage migration inhibitory factor	unfeasible	no	0	0	0
P00811	beta-lactamase	unfeasible	no	0	0	0
P11838	endothiapepsin	unfeasible	no	0	0	0
A9JQL9	dehydrosqualene synthase	unfeasible	no	0	0	0
P24627	lactotransferrin	unfeasible	no	0	0	0
P59071	phospholipase A2	unfeasible	no	0	0	0

[*]Successes are highlighted in green, while failures are highlighted in red. The number of times the program succeeded in the 10 jobs run is also indicated. [a]Selection of the solution with the best *AlignScore*. [b]Selection of the first ranked solution based on Borda tally. [c]Selection of the solution with the lowest strain energy.

scored better on at least one objective and did not score worse on any objective.

Only nondominated solutions were considered and subjected to the *AlignScore* analysis. In a real medicinal chemistry project, the user could manually inspect all solutions or make a decision on which to choose based on alternative analyses (e.g., clustering similar solutions based on common pharmacophores, good shape matching, or other approaches that consider known experimental preconditions). In the present work, we needed more objective rules for selecting only one nondominated solution as a result of each search. Therefore, we investigated the performance of the program in three separate experiments by choosing different types of solutions: (1) the solution returning the best *AlignScore* (this is, of course, an analysis only possible in retrospect because in real drug design projects the experimental overlay will not be available), (2) the highest ranked solution based on Borda tally, and (3) the solution with the lowest total internal strain energy calculated from the torsional and van der Waals terms of the Tripos force field.[15]

For each experiment we also evaluated the repeatability of results across 10 separate jobs (Tables S1−S3, Supporting Information).

RESULTS AND DISCUSSION

We tested the ability of the overlay program to reproduce the experimental overlay of 121 systems grouped into four categories (i.e., easy, moderate, hard, and unfeasible) on the basis of how easy or difficult it would be for a program to reproduce them. As a reminder, the validation protocol was started by generating ensembles of conformers for each set of ligands in the AZ test set using an in-house approach whose results are optimized to reflect the geometrical preferences of molecules observed in the CSD. The conformer generator was able to find at least one conformation with RMSD < 1.5 Å with

respect to the PDB structure for 96% of the ligands in the entire AZ test set (Figure S2, Supporting Information). The overlay program was run 10 times per set in order to avoid biases from the stochastic nature of the search algorithm. Each run resulted in up to 20 overlay predictions, among which only nondominated solutions were considered and subjected to the *AlignScore* analysis. In theory, they were all potential overlay hypotheses to test, but for validation purposes, it was necessary to adopt a consistent selection for all sets of ligands. Here, we analyzed the performance of the overlay program when using two prospective criteria of selection, and we compared these results with those obtained if only considering the solution having the best *AlignScore* (retrospective approach).

The results of the performance evaluation are presented in Table 2. Each set was deemed successful if the program satisfied the *AlignScore* criterion in at least one of the 10 jobs. The average repeatability for the four categories is shown in Table 3.

Table 2. Performance of Overlay Program against AZ Test Set

selected predicted overlay	no. of sets[a]	% of success			
		easy	moderate	hard	unfeasible
best *AlignScore*	121	95	73	39	0
	99	100	71	39	0
first-ranked solution	121	91	53	28	0
	99	80	51	28	0
lowest strain energy	121	95	49	28	0
	99	100	47	28	0

[a]121 ligand sets if considering the entire AZ test set; 99 sets when excluding those with at least 50% of congeneric ligands.

Table 3. Repeatability of Results across 10 Separate Job Runs

selected predicted overlay	average number of successful jobs			
	easy	moderate	hard	unfeasible
best *AlignScore*	9	6	3	0
first-ranked solution	8	3	2	0
lowest strain energy	8	3	1	0

When the entire set of 121 molecular overlays was considered, at best the overlay program was able to generate correct predictions for 95%, 73%, and 39% of the easy, moderate, and hard systems, respectively. None of the unfeasible sets was predicted correctly.

We found that the calculation of *AlignScore* was prohibitively time consuming for the target P00742 (i.e., coagulation factor XA, 36 ligands), and therefore, we could not evaluate how the program performed in this case. Nevertheless, it has been included in the results as a failure.

In addition to the best *AlignScore*, Table 2 also shows the results obtained in a prospective test where two different criteria for selecting an optimal prediction were adopted. In the first case, for each calculation, the top-ranked solution based on the Borda tallies of the objective functions was considered. In the second case, the solution having the lowest internal strain energy was considered. The great majority of the easy sets were predicted correctly regardless of the selection criterion considered. Also, in both experiments, in half of the moderate sets and 28% of the hard sets, the software reproduced the experimental superimposition of ligands.

One could argue that the *AlignScore* metric smoothed the distinction of easy and difficult systems when less than half the ligands within a set contributed to making it challenging. In these cases, the program could be successful (correctly aligning at least half of the ligands) by generating the right prediction only for the ligands that were similar to each other while failing to predict the more challenging diverse ligands. Therefore, to avoid any bias from the metrics used, we recalculated the percentages of success after excluding test sets where at least half of the ligands were part of the same congeneric series (results are shown in the bottom line of each experiment in Table 2). To do this, we identified the ligands in a set that had Tanimoto similarity values higher than 0.7 with at least 50% of the other ligands. Sets where at least half of the ligands met this criterion were considered congeneric and were excluded. This resulted in the exclusion of 17 of the 22 easy sets and 5 of the 73 moderate sets (Table 1) and gave success rates of 100%, 71%, and 39% for the easy, moderate, and hard systems, respectively.

The reason for the surprising increase in the success rate of the easy sets was due to the exclusion of the P54760 (ephrin type-B receptor 4) set for which the program was only able to reproduce the experimental overlay for four out of nine ligands.

Illustrative examples for each category will be discussed in the remainder of this section.

Discussion of Selected Examples. *Easy Set – P08235: Mineralocorticoid Receptor.* The overlay program generated the correct superimposition of all six ligands in this set. Figure 2 shows the predicted overlay with the best *AlignScore* (B) compared to the experimental overlay (A). The true answer was unambiguously reproduced for both steroidal and non-steroidal binders.

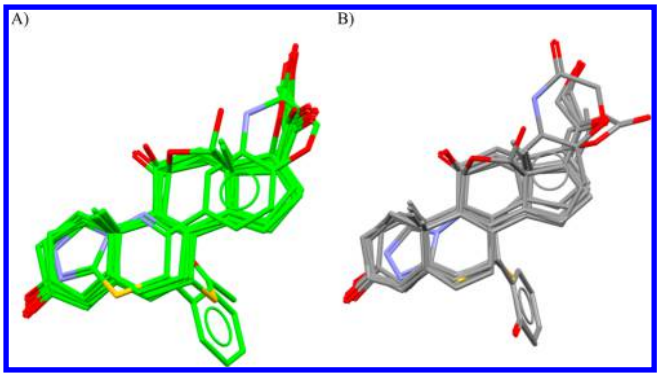


Figure 2. Comparison of the experimental (A) and predicted (B) overlays for the six ligands of the P08235 (mineralocorticoid receptor) set.

Moderate Set – P00749: Urokinase. This large set contained 27 ligands. The *AlignScore* for the best predicted overlay was 82%, which means that 22 (3kgp, 1owe, 2vin, 1gi7, 2vio, 2viw, 2viq, 2vip, 1sqt, 2viv, 1owd, 1sqo, 1gi8, 1gj7, 1c5y, 1gi9, 1o3p, 1gjc, 1o5a, 1o5c, 1gj8, 1gj9) ligands where overlaid in a way that reproduced the experimental observation (Figure 3). Figure 4 shows a tiled comparison between the "true" and the predicted conformation of the five failing molecules. For two of them (1u6q and 2vnt) the program was able to map the key features but positioned the solvent accessible chain in a different orientation to that of the protein–ligand complex. A

dx.doi.org/10.1021/ci500509y | J. Chem. Inf. Model. 2014, 54, 3091–3098

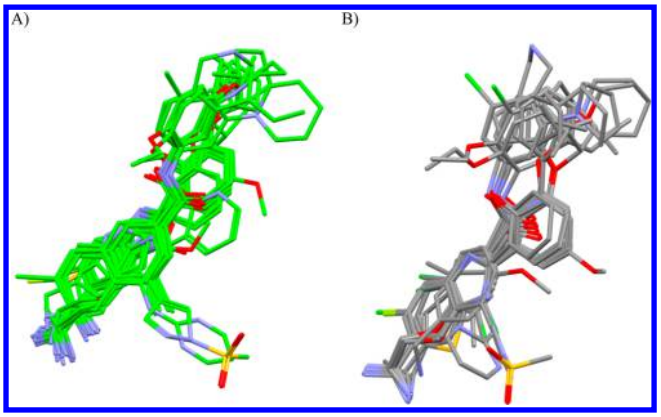


Figure 3. Comparison of the experimental (A) and predicted (B) overlays for the 22 ligands of the P00749 (urokinase) set, which contributed to the 82% success rate.

reasonable prediction was obtained for the molecule 3mwi albeit with an RMSD higher than the threshold.

Hard Set – P68400: Casein Kinase II. At best, the program generated a correct prediction for eight (3axw, 3owk, 3mb7, 3at4, 3pe1, 3r0t, 3pe2, 3mb6) of the 14 ligands in this set, returning an *AlignScore* of 57%. The predicted overlay for these structures is shown in Figure 5A and B in comparison with their equivalents in the experimental overlay. The program did not reproduce the experimental conformation of six ligands. However, they shared only the aromatic features with the remaining molecules in the set; none of them contained the

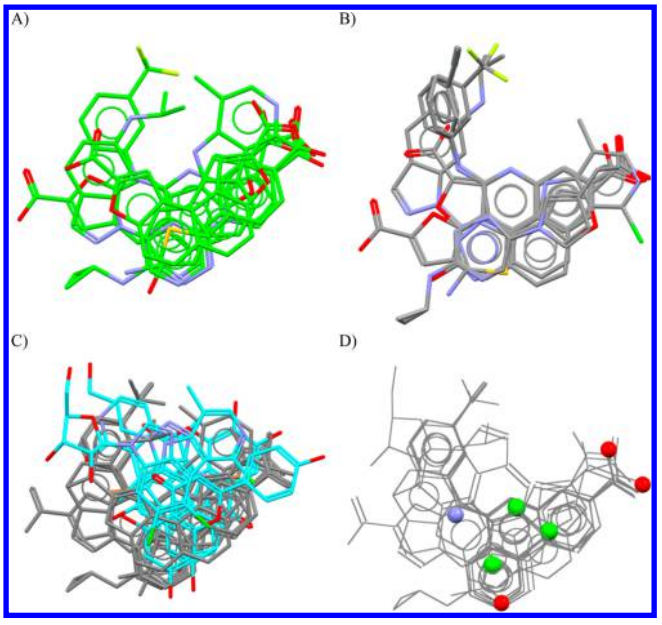


Figure 5. Comparison of experimental (A) and predicted (B) overlays for eight of the 14 ligands in the P68400 (casein kinase II) set. (C) Experimental overlay of the six misaligned ligands (cyan) superimposed onto the successful molecules (gray). (D) Key pharmacophoric features observed in the "true" overlay as generated by the CSD-driven overlay program are represented as spheres. Aromatic, donor, and acceptor points are colored in green, blue, and red, respectively.

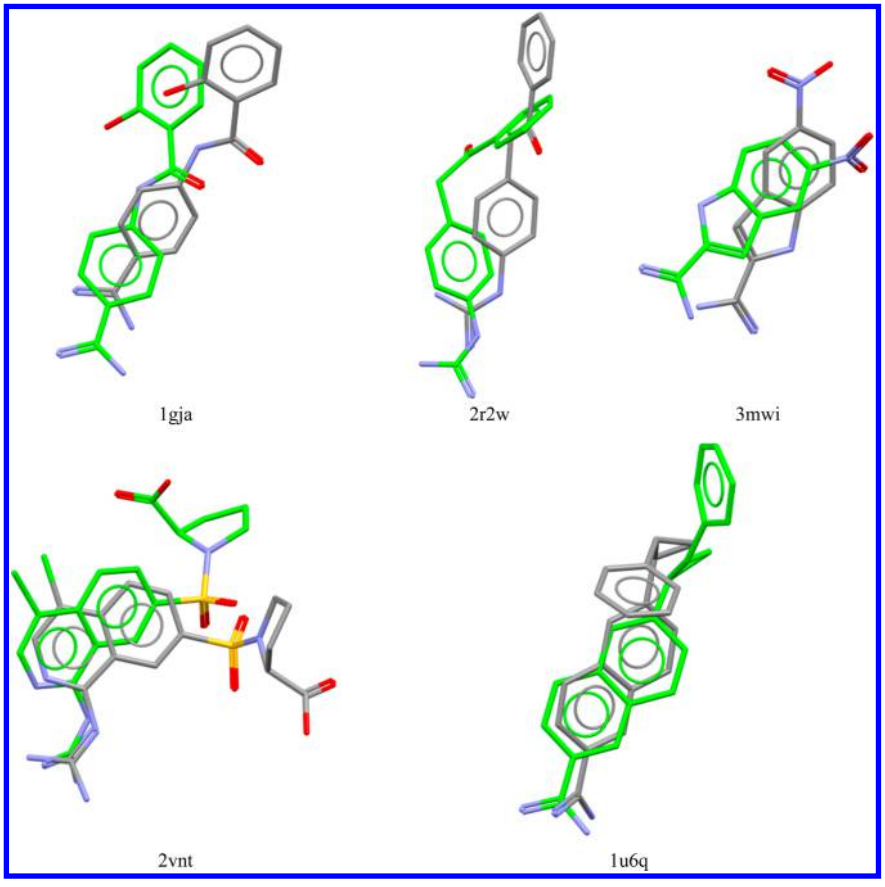


Figure 4. Experimental conformation (green) and best predicted solution (gray) of the five misaligned ligands of the P00749 (urokinase) set. The PDB code of each ligand is shown.

dx.doi.org/10.1021/ci500509y | *J. Chem. Inf. Model.* 2014, 54, 3091−3098

acceptor and/or the donor points. The observed poses of ligands 3amy, 3owj, and 3owl even contained a donor feature superimposed onto the acceptor point highlighted in Figure 5D, demonstrating the inherent difficulties of overlaying ligands based on their chemical features. Finally, ligands 3h30 and 3rps were bound to the interdomain hinge region of the protein through a weak halogen bond, which the program does not currently take into account.

Unfeasible Sets. Our classification approach led to the conclusion that it would be unfeasible to predict the experimental overlay of eight sets, and "correct" predictions would be chance events. Visual inspection of these sets reinforced their unfeasibility. For example, dehydrosqualene synthase (UniProt ID: A9JQL9) contains a structure (i.e., 2zcp) with two copies of the same ligand in the same cavity.

The caspase-3 (UniProt ID: P42574) example contains seven ligands observed as adducts with the catalytic cysteine but with no clear pharmacophoric feature common to all or some of them. Although the experimental overlay can suggest the shape of the pocket accessible for ligand binding, some molecules of this set recognize only a small region of the binding site. A partial overlay is then observed, and it is unlikely to be predicted computationally in the absence of additional binding information.

Ten of the 11 structures in the endothiapepsin set (UniProt ID: P11838) are discussed in the same work by Koster et al.[16] They adopted a fragment-based approach to find endothiapepsin inhibitors and determined crystal structures with a number of fragments. Different binding modes were observed, mainly distinguished into three categories: (1) direct binding to the two catalytic aspartates, (2) binding to the two catalytic aspartates mediated by a water molecule, and (3) no direct interaction with the catalytic dyad. However, even ligands belonging to the same class appeared to have no overlap when bound to the protein. Such an arrangement (Figure 6) is impossible to determine without structural information.

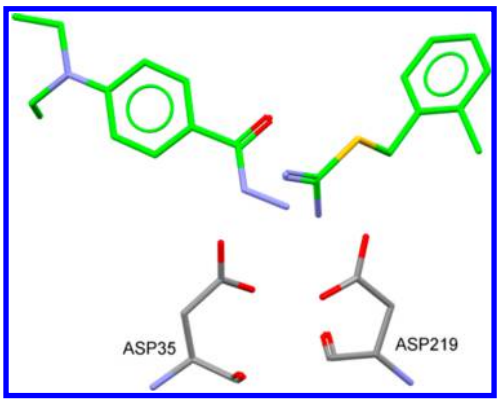


Figure 6. Two endothiapepsin inhibitors are shown, rendered as green sticks. They share a direct binding mode to the two catalytic aspartates (Asp35 and Asp219) of the protein but occupy different regions of the binding site.

Belonging to our unfeasible class, lactoferrin (UniProt ID: P24627) is a multifunctional protein that mediates various functions. Its promiscuity in binding different types of ligands underlines the absence of shared pharmacophoric features between the ligands in this set. Similar considerations can be drawn for the remaining four unfeasible sets.

One can regard the unfeasible set as a negative control on the underlying results. It would be unrealistic to expect a method that assumes a pharamcophore hypothesis to be able to predict structural results that do not conform to said hypothesis. If they did so, one could conclude that certain correct results could have been generated by chance. The result that this program never generates a correct answer for this set, then, gives confidence that the overlays generated have a sound underlying basis.

CONCLUSIONS

We have performed an extensive assessment study of a recently developed algorithm to overlay multiple flexible ligands. We tested the ability of the program to reproduce the experimental overlay of 121 ligand sets available from the recently published AZ test set. In order to provide a baseline validation, default settings were adopted, and repeatability of the results was verified by running the same job 10 times for each ligand set. In addition, three criteria were used to select only one optimal overlay hypothesis from the pool of nondominated solutions. As in previous works (i.e., validation of GAPE and FLAPpharm), we compared a given predicted overlay with the one derived experimentally by using a metric called *AlignScore*, which represents the percentage of ligands in the predicted superimposition that can be fitted to the crystal structure alignment with a heavy atom RMSD $\leq$ 2 Å. The prediction was deemed successful when *AlignScore* was equal to or better than 50%. We assessed the performance of the overlay program based on the level of difficulty of the analyzed sets. To this aim, the AZ test set was reranked and categorized into four classes. We found success rates of 95% for the easy sets, 73% for the moderate sets, and 39% for the hard sets. At these levels, the repeatability of success was 9, 6, and 3 out of 10 jobs, respectively.

As existing software for pharmacophore elucidation and fitting has not been assessed using these objective metrics against such a large range of systems, it is not possible for us to compare the relative performance of the CSD-driven overlay tool. This software is, however, already in use in a number of institutions, and we look forward to further evaluation.

However, we can say at this point that the success and robustness of the software do correlate with the complexity of the system under investigation.

Certainly, if enough information is available in the chemical features of the ligands, this program will produce a set of plausible overlays that can, in turn, be validated by protein–ligand crystal structure determination and directed medicinal chemistry.

ASSOCIATED CONTENT

Supporting Information

Description of the *AlignScore* algorithm, example of *AlignScore* evaluation, *AlignScore* values for all experiments done, and assessment of the conformer generation. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*E-mail: giangreco@ccdc.cam.ac.uk.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank Dr. Robin Taylor and Dr. Colin Groom for their reviews of this manuscript.

■ REFERENCES

(1) Leach, A. R.; Gillet, V. J.; Lewis, R. A.; Taylor, R. Three-dimensional pharmacophore methods in drug discovery. *J. Med. Chem.* **2010**, *53*, 539−558.

(2) Miettinen, K. *Nonlinear Multiobjective Optimization*; Springer: New York, 1999.

(3) The Borda tally was introduced in the 18th century as a single-winner election method. Each voter provides a ranked ordering of all candidates, which are assigned with a certain number of points based on their position. Once all votes have been counted, the candidate with the most points is the winner. In the context of this work, the Borda tally is the sum of the ranks of the individual objective scores.

(4) Giangreco, I.; Cosgrove, D. A.; Packer, M. J. An extensive and diverse set of molecular overlays for the validation of pharmacophore programs. *J. Chem. Inf. Model.* **2013**, *53*, 852−866.

(5) Taylor, R.; Cole, J. C.; Cosgrove, D. A.; Gardiner, E. J.; Gillet, V. J.; Korb, O. Development and validation of an improved algorithm for overlaying flexible molecules. *J. Comput.-Aided Mol. Des.* **2012**, *26*, 451−472.

(6) Rush, T. S.; Grant, J. A.; Mosyak, L.; Nicholls, A. A shape-based 3-D scaffold hopping method and its application to a bacterial protein−protein interaction. *J. Med. Chem.* **2005**, *48*, 1489−1495.

(7) OpenEye Scientific Software OEShape Toolkit. http://www.eyesopen.com/toolkits.

(8) Blomberg, N.; Cosgrove, D. A.; Kenny, P. W.; Kolmodin, K. Design of compound libraries for fragment screening. *J. Comput.-Aided Mol. Des.* **2009**, *23*, 513−525.

(9) Sadowski, J.; Gasteiger, J.; Klebe, G. Comparison of automatic three-dimensional model builders using 639 X-ray structures. *J. Chem. Inf. Model.* **1994**, *34*, 1000−1008.

(10) Allen, F. H. The Cambridge Structural Database: A quarter of a million crystal structures and rising. *Acta Crystallogr., Sect. B: Struct. Sci.* **2002**, *58*, 380−388.

(11) Bruno, I. J.; Cole, J. C.; Kessler, M.; Luo, J.; Motherwell, W. D. S.; Purkis, L. H.; Smith, B. R.; Taylor, R.; Cooper, R. I.; Ox, O.; Harris, S. E.; Orpen, A. G. Retrieval of crystallographically-derived molecular geometry information. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 2133−2144.

(12) Taylor, R.; Cole, J.; Korb, O.; McCabe, P. Knowledge-based libraries for predicting the geometric preferences of druglike molecules. *J. Chem. Inf. Model.* **2014**, *54*, 2500−2514.

(13) 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 data set for evaluating pharmacophore elucidation methods. *J. Chem. Inf. Model.* **2012**, *52*, 2599−2608.

(14) Jones, G. GAPE: An improved genetic algorithm for pharmacophore elucidation. *J. Chem. Inf. Model.* **2010**, *50*, 2001−2018.

(15) Clark, M.; Cramer, R. D.; Van Opdenbosch, N. Validation of the general purpose Tripos 5.2 force field. *J. Comput. Chem.* **1989**, *10*, 982−1012.

(16) Köster, H.; Craan, T.; Brass, S.; Herhaus, C.; Zentgraf, M.; Neumann, L.; Heine, A.; Klebe, G. A small nonrule of 3 compatible fragment library provides high hit rate of endothiapepsin crystal structures with various fragment chemotypes. *J. Med. Chem.* **2011**, *54*, 7784−7796.