Geometric Accuracy of Three-Dimensional Molecular Overlays

Qi Chen,* Richard E. Higgs, and Michal Vieth

Lilly Research Laboratories, Eli Lilly and Company, Lilly Corporate Center, DC 1930, Indianapolis, Indiana 46285

Received April 11, 2006

This study examines the dependence of molecular alignment accuracy on a variety of factors including the choice of molecular template, alignment method, conformational flexibility, and type of protein target. We used eight test systems for which X-ray data on 145 ligand−protein complexes were available. The use of X-ray structures allowed an unambiguous assignment of bioactive overlays for each compound set. The alignment accuracy depended on multiple factors and ranged from 6% for flexible overlays to 73% for X-ray rigid overlays, when the conformation of the template ligand came from X-ray structures. The dependence of the overlay accuracy on the choice of templates and molecules to be aligned was found to be the most significant factor in six and seven of the eight ligand−protein complex data sets, respectively. While finding little preference for the overlay method, we observed that the introduction of molecule flexibility resulted in a decrease of overlay accuracy in 50% of the cases. We derived rules to maximize the accuracy of alignment, leading to a more than 2-fold improvement in accuracy (from 19% to 48%). The rules also allowed the identification of compounds with a low (<5%) chance to be correctly aligned. Last, the accuracy of the alignment derived without any utilization of X-ray conformers varied from <1% for the human immunodeficiency virus data set to 53% for the trypsin data set. We found that the accuracy was directly proportional to the product of the overlay accuracy from the templates in their bioactive conformations and the chance of obtaining the correct bioactive conformation of the templates. This study generates a much needed benchmark for the expectations of molecular alignment accuracy and shows appropriate usages and best practices to maximize hypothesis generation success.

INTRODUCTION

Structurally different molecules often bind in the same binding sites and establish functional group equivalencies. Understanding the spatial relationship of functional groups of different molecules is necessary to develop three-dimensional (3D) pharmacophore hypotheses. Such hypotheses have been utilized in various aspects of drug design, such as structure−activity-relationship (SAR) rationalization and design, in silico screening, and scaffold hopping.[1−9] The development of a three-dimensional pharmacophore model involves the generation of conformations of molecules and overlaying them using algorithms to maximize an objective function such as shape, functional group, and property distribution overlap. There are a number of commercially available software packages which perform this task, including FlexS,[10] ROCS,[11] Catalyst,[2] Seal,[12] and MIMIC.[13]

To ensure that results are meaningful, one needs to perform a proper validation of this structural hypothesis. A proper structural validation of a 3D pharmacophore hypothesis is, however, not a trivial task in the absence of an X-ray crystal structure of the drug−protein complex. This is unfortunately the case for most therapeutically relevant protein targets, including G-protein coupled receptors. In numerous instances, researchers have used indirect validation procedures including the agreement between their intuition, SAR trends, 3D quantitative structure−activity relationships, and the overlay results.[3,4,8,14]

A compelling example of the direct structural validation of 3D overlays included the utilization of 76 X-ray structures of complexes from 14 proteins with the FlexS program.[10] The number of compounds in each protein data set varied from 2 to 12. This work used the experimentally determined X-ray complexes to judge the quality of the overlay results. The results showed that for more than 40% of the molecule−template pairs the correct overlays could be found among the 10 best-scored overlays from each aligned molecule. The X-ray conformations of the compounds were used as templates in this study.

The general outcome of 3D overlays can be assumed to depend on the choice of the template molecule, the conformational difference of the template with respect to the bioactive conformation (i.e., the protein-bound conformation from the X-ray crystal structure), the conformational space of the molecules being overlaid, and the overlay method. The choice of the template and the molecules has not been thoroughly investigated. Most programs use the preselection of templates and sometimes the molecules being aligned to avoid a combinatorial explosion of the search space. For example, Catalyst[2] uses the potency range to select only a few compounds as the templates from which a pool of possible models are generated. ROCS,[11] FlexS,[10] and Seal[12] require users to provide the initial 3D structure of a template molecule.

In this work, we examine multiple factors influencing the quality of overlay results from two of the most commonly used packages: ROCS and FlexS. Following docking studies

* Corresponding author phone: 317-277-3016; fax: 317-276-6545; e-mail: qchen@lilly.com.

10.1021/ci060134h CCC: $33.50
Published on Web 07/11/2006

Table 1. Overlay Options for the Molecules Being Aligned[a]

	ROCS−rigid	ROCS−flexible	FlexS−rigid	FlexS−flexible
overlay type	single conf.	multiple conf.	rigid	flexible
conformation	X-ray	OMEGA	X-ray	CORINA

[a] The options for ROCS were -chemff MillsDean.cff, -optchem, and -rankby combo. The default conditions were used in FlexS, including a 60% volume overlap threshold. The options for OMEGA were up to 500 conformations saved, an energy window of 6 kcal/mol, and a RMSD for duplicates of 0.5 Å, using CORINA conformations as input.[21] The average number of conformations per compound was 233 with a standard deviation of 219. The CORINA conformations used the unique SMILES as input.

which also produce molecular overlays, we utilize commonly accepted root-mean-squared-deviation (RMSD) metrics to judge the structural accuracy of molecular overlays.[15] Overlaid X-ray complexes of 145 ligands in eight proteins were used. We examine the dependence of the overlay results on template choice, template conformation, overlay method, input conformations of the overlaid molecules, and the protein target test system. The results are then used to formulate guidelines for the best overlay practices for the selection of templates and molecules. From the results, we also propose a hypothesis that the accuracy of molecular overlays can be decomposed into a template conformation component and overlay accuracy component given the bioactive-conformation-like template. Finally, the results offer a realistic expectation of the overlay accuracy for cases where X-ray crystal structures are not available. The overlay accuracy is further compared with the docking accuracy reported by Erickson et al.[15]

METHODS

Molecular Overlay Data Sets. We have utilized X-ray complexes[16] of 145 unique compounds with eight protein targets (cyclin-dependent kinase 2, CDK2; human immunodeficiency virus, HIV; estrogen receptor 1, ESR1; mitogen-activated protein kinase 14, p38; thermolysin; rhinovirus; elastase; and trypsin) to perform testing of the molecular overlay accuracy. The number of compounds for each target ranged from 7 (trypsin and elastase) to 57 (CDK2). For each of eight protein targets, the corresponding compounds were aligned inside their protein-binding sites on the basis of the alignment of protein structures. We will refer to the structures from such an alignment as the X-ray-based aligned structures in order to distinguish them from those using the ligand-based alignment methods. The aligned structures of 111 compounds for CDK2, HIV, ESR1, and p38 were obtained through the protocol described previously,[15,17] which is virtually equivalent to Quanta[18] alignments. The aligned structures of 34 compounds for elastase, trypsin, thermolysin, and rhinovirus were obtained from the FlexS-77 data sets.[10] These eight sets of the X-ray-based aligned structures provided experimentally determined bioactive conformations and served as the reference for assessing the geometric quality of the ligand-based alignment.

Molecular Alignment Procedure. For each of eight data sets, every ligand from the X-ray crystal structure was used as the template. ROCS (v.2.1.1) and FlexS (v.1.11.4) molecular overlay tools were applied to align all molecules (in their X-ray conformations and in silico conformations) to the template. Only the best scoring overlay for a molecule−template pair was selected for consideration. The heavy atom positions in the computationally aligned molecules were compared to those in the X-ray-based aligned structures. Symmetry considerations were taken into account in the calculated RMSD values[15] for each molecule. The overlays with heavy atom RMSD values less than or equal to 2 Å from X-ray positions were considered correct. The procedure resulted in the generation of 4677 RMSD values for 145 templates from every overlay method (Table S2 of the Supporting Information). A total of 57 CDK2 molecules accounted for 3249 molecule−template pairs (39% of the molecules and 69% of the pairs). Elastase and trypsin with seven molecules per set accounted for 49 molecule−template pairs. The overlay accuracy for a template was defined as the percentage of the correct overlays (corr%). Arguably, the heavy atom RMSD is one of many possible metrics to judge the quality of molecular overlays, and we have selected it for simplicity and to establish a direct comparison to other techniques such as molecular docking. Our conclusions on the importance of different factors need to be interpreted with the metrics in mind.

Conformational Flexibility. We considered three means of establishing initial conformations for each compound: the protein-bound structures from the X-ray complex and computationally generated conformations—a single conformation from CORINA[19] (from unique SMILES) and multiple conformations from OMEGA (v.1.8.1)[20] using CORINA as input.[21]

While both ROCS and FlexS hold template conformations rigid during overlays, they use different ways to handle the flexibility of the molecules to be aligned. ROCS utilizes multiple conformations of a molecule and rigidly aligns each of them to the template. FlexS takes a single conformation of each molecule and changes the conformation during the alignment protocol. For each molecule−template pair, we tested the performance of four different overlays: two overlay methods—each in rigid and flexible modes, as summarized in Table 1. X-ray conformations of compounds served as templates in this part of the study.

Because in real drug discovery problems the protein-bound conformations of the templates are often not known, we have also used computationally (CORINA) derived conformations for the templates. The use of CORINA conformations allowed us to investigate the effect of structural deviations of a template on overlay accuracy. In this final analysis, the conformations of both templates and molecules being aligned were generated through computational means. In our opinion, this provided estimates for a more realistic overlay accuracy.

Molecular Descriptors. For each molecule−template pair, we have calculated the Tanimoto similarity[22,23] of Daylight 2048 fixed-width fingerprints[24] (SimDY) and the Tanimoto similarity of the number of heavy atoms (SimNA). In addition, the number of rotatable bonds (#rotB), hydrogen-bond donors and acceptors, computed logP, polar surface area, atomic molecular weight, and ratios of these descriptors

between the molecules and templates were calculated.[25] The mean values over all molecules in the same data set have been computed for each template (e.g., AveSimDY for mean Tanimoto fingerprint similarity). A summary of these molecular descriptors and similarities is shown in Table S1 of the Supporting Information.

Significance Analysis and Rules. We have used the effect test of a logistic regression[26] to examine the significance of the differences of molecular overlay accuracy as a function of methods (ROCS vs FlexS), flexibility handling (rigid vs flexible overlay), protein targets (eight data sets), and templates and molecules being overlaid (per target). The response in the model fitting was the correctness of the overlay, that is, correct or incorrect. There were 18 708 data points in total from four runs as depicted in Table 1. A χ-squared test p value less than 0.05 for a specific variable indicates that the factor plays a statistically significant role in determining the overlay accuracy. The analyses were performed using the JMP 5.1.1 software.[27]

Recursive partitioning[28] modeling in JMP was used to explore the relative importance of various molecular descriptors in the overlay accuracy and to develop empirical rules to maximize the accuracy of overlay results. The CDK2 data set was used as the training set to derive the rules. Each rule was then applied to other targets to validate the applicability of the rules.

RESULTS AND DISCUSSION

We report the analysis of molecular overlays in three sections. First, we focus on analyzing the effects of possible factors which influence molecular overlay accuracy. We then use the lessons from this analysis and provide guidance on how to maximize the probability of achieving a correct overlay. Finally, we report the realistic benchmark results assuming that all molecules bind to the same binding site but no explicit knowledge of the molecules, target, or conformations is known. To provide a visual representation of our work, we have included structures of a CDK2 ligand 1ogu aligned on various templates in comparison to its bioactive conformation in Figure 1. This figure exemplifies the template dependence of the alignment with correct ($\leq$2 Å) and incorrect (>2 Å) overlays.

1. Effects on the Overlay Accuracy. We found significant and meaningful dependence of the overlay accuracy results on the choice of the template and aligned molecules in 80% of the cases and on the overlay type (rigid vs flexible) in 50% of the cases (Tables 2 and 3). We found little evidence that one overlay method is better than the other, as two different methods (ROCS and FlexS) resulted in similar performance. Therefore, we limited the discussion to issues common to the different methods, instead of the methodologies themselves. The detailed discussion of each factor follows in the subsequent sections.

Templates And Molecules. The alignment results were very sensitive to the choice of templates, regardless of the methods and conformational choices for seven of the eight test system proteins (Table 2 and Table S2 of the Supporting Information). For example, the best CDK2 template (1oiy) for ROCS−flexible gave 25 correct overlays, while the worst template (1p5e) had only one. The difference in the accuracy from different templates varied up to 72% (e.g., trypsin) for

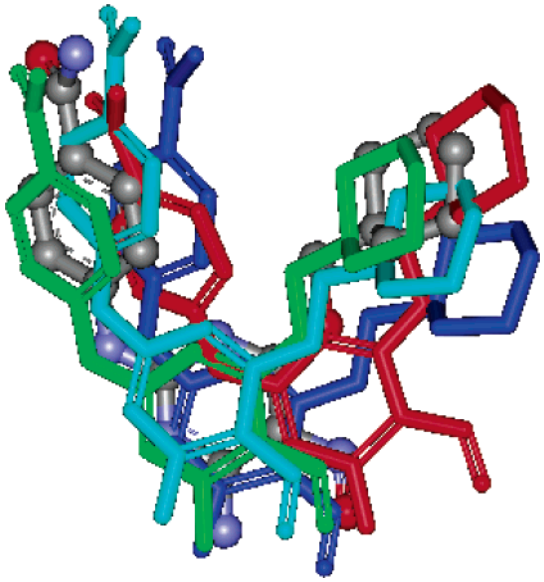

Figure 1. Aligned structures of the CDK2 ligand 1ogu aligned by the superposition of protein complexes (ball-and-stick, gray) and by ROCS using the following templates: 1urw (RMSD = 1.09 Å, green), 1pxl (1.47 Å, cyan), 1ke7 (1.82 Å, blue), and 1gih (2.27 Å, red), where the RMSD value refers to the deviation between the ROCS-aligned 1ogu structure and its bioactive conformation (i.e., ball-and-stick). The overlap of the functional groups in these aligned structures with those in the bioactive conformation is reasonably good, except for the overlaid structure with RMSD = 2.27 Å. Those showing good overlap have RMSDs from 1.09 to 1.82 Å, consistent with the commonly used 2 Å RMSD cutoff. The graphic was produced from Weblab Viewerlite 4.0 (Accelrys Inc. Catalyst v.4.9; Accelrys Inc.: San Diego, CA).

Table 2. Alignment Accuracy (corr%) from the Best, Worst, and Average of All Templates in Their Bioactive Conformations[a]

		rigid (%)			flexible (%)		
protein[b]	method	best	worst	average	best	worst	average
CDK2 (57)	ROCS	63	2	30	44	2	20
	FlexS	47	2	25	46	2	21
HIV (28)	ROCS	68	4	39	21	0	6
	FlexS	54	0	24	25	0	8
p38 (13)	ROCS	54	8	27	38	8	22
	FlexS	46	8	27	38	0	24
ESR1 (13)	ROCS	54	23	44	46	8	25
	FlexS	85	15	47	38	0	28
elastase (7)	ROCS	43	14	31	43	0	20
	FlexS	57	0	31	43	14	27
trypsin (7)	ROCS	86	14	57	71	14	55
	FlexS	86	57	73	86	0	29
thermolysin (12)	ROCS	33	8	19	25	0	10
	FlexS	50	17	30	42	8	19
rhinovirus (8)	ROCS	50	50	50	50	50	50
	FlexS	63	50	52	50	50	50
total (145)	ROCS	59	9	34	39	6	20
	FlexS	55	10	32	42	5	21

[a] For more details, see Table S2 in the Supporting Information. [b] The number in parentheses is the number of ligands in the data set.

the same set of compounds. On average, the accuracy for the best templates was 33−50% better than that of the worst templates across the four overlay runs. The rhinovirus system did not show any template dependence, possibly because the compounds in this set were highly similar. For this data set, the mean AveSimDY was 0.94, compared to only 0.27−0.58 for the seven other systems. It was also reported that they had two different binding modes,[10] leading to an overall 50% overlay accuracy, which was lower than expected for

Table 3. Statistical Test: *P* value and the Odds Ratios (in Parentheses) from the Effect Test of the Logistic Regression Fitting Models[a]

			rigid vs flexible		ROCS vs FlexS	
protein	template	molecule	ROCS	FlexS	rigid	flexible
CDK2	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.0001**	0.416
			(1.99)	**(1.39)**	**(1.33)**	(0.94)
HIV	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**0.039**
			(15.93)	**(5.43)**	**(2.70)**	**(0.65)**
p38	**0.023**	**<0.001**	0.273	0.410	0.894	0.674
			(1.35)	(1.25)	(0.97)	(0.89)
ESR1	**0.007**	**<0.001**	**<0.001**	**<0.001**	0.481	0.546
			(2.46)	**(3.24)**	(0.85)	(0.83)
trypsin	**0.000**	**<0.001**	0.757	**<0.001**	**0.028**	**0.002**
			(1.21)	**(27.21)**	**(0.25)**	**(6.65)**
thermolysin	0.266	**<0.001**	**0.026**	**0.022**	**0.020**	**0.015**
			(2.26)	**(2.06)**	**(0.50)**	**(0.40)**
elastase	**0.009**	**0.014**	0.207	0.612	1.000	0.421
			(1.93)	(1.29)	(1.00)	(0.65)
rhinovirus	1.000	1.000	1.000	0.859	0.859	1.000
			(1.00)	(1.06)	(0.94)	(1.00)

[a] The formula of the models was $Y = f$(template, molecule, method), where Y was either the correct or incorrect overlay. Separate models were built for each target and each pair of method comparisons, i.e., ROCS−rigid vs ROCS−flexible, FlexS−rigid vs FlexS−flexible, ROCS−rigid vs FlexS−rigid, and ROCS−flexible vs FlexS−flexible. The numbers in the parentheses are the odds ratios from rigid vs flexible overlays, or from ROCS vs FlexS. The bold-text cells indicate that the differences are significant with a P value < 0.05, and thus, the overlay accuracy depends on the particular factor. For significant differences, the odds ratio is a way of comparing whether the probability of a certain event is the same for two groups. An odds ratio of 1 implies that the event is equally likely in both groups. An odds ratio greater than 1 implies that the event is more likely in the first group. An odds ratio less than 1 implies that the event is less likely in the first group.

such highly similar compounds. The effect test results from the logistic regression modeling confirmed that the dependence of the overlay accuracy on the template choice was statistically significant for six of the eight test systems (Table 3). The dependence on the overlaid molecules was also significant for all proteins except rhinovirus (Table 3). For example, CDK2 molecule 1ogu was correctly aligned with 30 templates in flexible ROCS, while 1p5e was correctly aligned by only four templates.

Flexibility of the Compounds. Seven of eight targets (except rhinovirus) exhibited lower accuracy in the flexible overlay than in the rigid overlay. The effect test suggested that the accuracy reductions were statistically significant for five test systems including CDK2, HIV, ESR1, trypsin (FlexS only), and thermolysin. The largest difference was observed for the most flexible HIV compounds, where the mean loss of accuracy was 33% for ROCS and 16% for FlexS. This relatively large loss of overlay accuracy is consistent with the fact that the average number of rotatable bonds was 16 for HIV compounds compared to 3−8 for the other targets. Consistent with this larger flexibility, only 4% of the HIV compounds had their CORINA-generated computational conformations within a 2 Å RMSD of the bioactive conformations. This is in contrast to other test systems for which 68% had CORINA conformations within a 2 Å RMSD. Up to 500 computationally generated OMEGA conformers improved the occurrence of bioactive-like conformers, but the trend remained as the HIV data set had 57% compounds with bioactive-like representatives versus 97% for all other compounds. The lack of a significant reduction in overlay accuracy was observed in more rigid systems such as p38 (mean of 3.5 rotatable bonds). Trypsin, the most rigid system, with a mean of 2.8 rotatable bonds, showed flexibility dependence only for FlexS. This difference originates from four templates which failed to align correctly any molecules in flexible overlays but were exceptionally good in rigid overlays with the accuracy ranging from 57% to 86%.

Overlay Algorithm. The differences in overlay accuracy between the different overlay methods ROCS and FlexS were small (Table 2) and significant in only 7 of 16 cases (four rigid and three flexible, Table 3). In these seven cases, the odds ratios indicated that FlexS was better than ROCS in two of three flexible cases and two of four rigid cases while ROCS prevailed in the remaining three cases (Table 3). The relative performance of a template was also very similar in the two methods. Overall, the squared correlation coefficient (R^2) of the overlay accuracy between FlexS−flexible and ROCS−flexible was 0.41 for all 145 templates. The CDK2 and elastase templates showed the highest R^2 values, 0.74 and 0.75, respectively. For the HIV test system, the R^2 was random for 28 templates. This was also the system with much lower overall accuracy likely due to highly flexible molecules. The significant difference observed in HIV was consistent with the significant P value of 0.039. On the basis of the small number of test systems where the differences between methods were significant and the fair correlation of the template performance in the two methods, our study did not indicate that one method/algorithm is better than the other. In terms of speed, flexible overlays with ROCS were significantly quicker with timing between 5 and 32 s on a 400 MHz R12000 CPU per molecule versus 1−4 min for flexible FlexS overlays.

2. Guidance to Enhance Overlay Accuracy. The detailed results from Tables 2 and 3 indicated that there is a significant dependence of the alignment accuracy on the template and molecule selection. On the basis of these results, we will address two practical issues: (1) how to utilize molecular descriptors to select compounds to maximize the probability of correct alignment, given a template, and (2) how to select the template to maximize the probability of correct alignment for a series of molecules.

The first issue addresses the ability to increase the reliability of scaffold hopping from selected seeds. The second relates to the development of a 3D pharmacophore hypothesis where the overlay of a series of compounds is required. We utilized various molecular properties to address these issues with flexible alignment. We applied JMP's recursive partitioning tool to establish relationships of descriptors to overlay accuracy. All analyses below are based on the results of flexible overlays obtained with templates in X-ray conformations.

Compound Selection: Similarity of Compound to Template (SimDY). SimDY was the most effective in differentiating the correct overlays from the incorrect ones. A total of 65% of the overlays were correct for the compounds with similarities to templates between 0.7 and 0.8, while for compounds with similarities to templates below 0.4, the correct overlays were below 30%. A large-scale study from 115 screening assays indicated that "there is only 30% chance that a compound that is ≥ 0.85 (Tanimoto) similar to an active is itself active".[29] Thus, 3D overlays might be useful in complementing the commonly used 2D approaches to identify additional protein ligands from not only compounds

Table 4. Guidance for Identifying Molecule−Template Pairs with a High or Low Chance of Being Correctly Aligned: Overlay Accuracy from the CDK2 Training Set

				ROCS−flexible		FlexS−flexible	
type	no.	rule description	total pairs	corr#	corr%	corr#	corr%
low chance	1	SimDY < 0.55 & #rotB ≥ 8	369	21	5.7%	13	3.5%
	2	SimDY < 0.24 & NAratio ≥ 1.28	596	24	4.0%	6	1.0%
	3	#1 or #2	874	44	5.0%	19	2.2%
	4	excluding pairs selected by #3	2373	612	25.8%	660	27.8%
high chance	5	(SimDY ≥ 0.24 & SimNA ≥ 0.74 & #rotB ≤ 11) − #3	768	376	49.0%	376	49.0%
	6	SimDY ≥ 0.55 & #rotB < 8	241	166	68.9%	175	72.6%
	7	SimDY ≥ 0.55 & #rotB = 8−11	30	18	60.0%	17	56.7%
benchmark		all molecule−template pairs	3249	656	20.2%	679	20.9%

Table 5. Guidance: Overlay Accuracy (corr%) from Seven Other Targets as the Test Sets[a]

rule no.[b]	elastase	ESR1	HIV	p38	rhinovirus	thermolysin	trypsin
				A. ROCS			
1	0.0	0.0	3.8	n.c.	n.c.	0.0	n.c.
2	0.0	0.0	0.0	2.3	n.c.	0.0	n.c.
3	0.0	0.0	3.8	2.3	n.c.	0.0	n.c.
4	23.8	34.4	17.9	28.8	50.0	16.3	62.8
5	36.0	50.9	42.9	57.1	50.0	35.0	73.0
6	50.0	54.2	n.c.	73.7	n.c.	52.6	75.0
7	60.0	n.c.	40.0	n.c.	50.0	60.0	n.c.
all pairs	20.4	25.4	5.7	21.9	50.0	9.7	55.1
				B. FlexS			
1	0.0	0.0	1.9	n.c.	n.c.	4.1	n.c.
2	0.0	2.8	0.0	2.3	n.c.	0.0	n.c.
3	0.0	2.3	1.9	2.3	n.c.	3.5	n.c.
4	31.0	36.8	49.1	31.2	50.0	30.2	32.6
5	44.0	49.1	28.6	67.4	50.0	30.0	32.4
6	66.7	50.0	n.c.	73.7	n.c.	63.2	**25.0**
7	60.0	n.c.	26.7	n.c.	50.0	40.0	n.c.
all pairs	26.5	27.8	8.3	23.7	50.0	19.4	28.6

[a] Each value of corr% was calculated from at least five molecule−template pairs. The average number of pairs per corr% value was 93. n.c. indicates where < 5 pairs were present. [b] See Table 4 for the definition of the rules that were derived from the CDK2 data set.

highly similar to an active but also those very structurally different from the active.

Compound Selection: Improving the Chance of Correct Alignment. SimDY, SimNA, and the number of atoms ratio (NAratio) of the molecule to the template, together with #rotB in the molecule, improved the chances of identification of correct as well as incorrect overlays. In addition, these descriptors allowed for the expansion of compounds to be successfully aligned to the broader regions of structural space.

Table 4 includes a set of rules initially derived from the CDK2 data set (training set) for compound selection to maximize the correct overlays. Each rule was then applied to seven other targets (test set) to ensure transferability to other data sets. In this study, we did not use the traditional statistical measures (e.g., r2, q2, etc.) to evaluate the goodness of the models because of the uneven distribution of data points in the targets. Our goal is to ensure that the rules we derived from one target would be qualitatively applicable to most other targets.

Transferability of the rules was confirmed by the results in Table 5. As expected, all pairs selected from negative rules (1, 2, and 3) had significantly lower accuracy than the average. At the same time, all but one of the target sets from positive rules (4, 5, 6, 7) gave higher overlay accuracy (Table 5). The only violation was observed for the trypsin data set selected by rule 6 for FlexS (Table 5B). In addition, there was no improvement observed from rhinovirus because none of its pairs belonged to rules 1−3.

Overall, almost half of the 1052 pairs (768 from CDK2) selected by rule 5 were correctly aligned, including molecules with low Daylight similarity (0.24) and a high number of rotatable bonds (up to 11), compared to about 19% from all 4677 pairs without selection (Tables 4 and 5). These compounds had SimNA ≥ 0.74 and either SimDY ≥ 0.55 with #rotB ≤ 11 or SimDY = 0.24−0.55 with #rotB < 8. The requirement of a high similarity of the number of heavy atoms to the template (SimNA ≥ 0.74) suggested that only the alignment of similar-sized molecules was possible. On the contrary, less than 5% of the 1711 pairs (874 from CDK2) selected by rule 3 could be correctly aligned to their templates. These molecules had either SimDY < 0.55 and #rotB > 8 or SimDY < 0.24 and NAratio ≥ 1.28, that is, too large, too flexible, and too dissimilar from the templates. Simply removing these molecules from the whole data set would improve the accuracy from 19% to 28%.

Comparing results from rules 1, 6, and 7 indicates that the similarity was more essential than the number of rotatable bonds in the molecules. This indirectly suggested both ROCS and FlexS could handle the flexibility of the molecules reasonably well. For highly flexible compounds with 8−11 rotatable bonds, it was possible to have about a 50% chance of correct alignment provided they were highly similar to the templates with Daylight similarity (SimDY ≥ 0.55, rule 7).

Our recommendations to maximize success are then as follows: (1) exclude the compounds with either (a) SimDY < 0.55 and #rotB > 8 (i.e., rule 1) or (b) SimDY < 0.24 and NAratio ≥ 1.28 (i.e., rule 2) and (2) include the compounds with SimNA ≥ 0.74 to the template and either (a) SimDY ≥ 0.55 to the template with #rotB ≤ 11 (rule 5) or (b) SimDY = 0.24−0.55 to the template with #rotB < 8 (rule 5).

Selecting the Optimal Templates. On the basis of the recursive partitioning results, we found that templates with higher average similarities to the compound set (δAveSimDY ≤ 0.06 and δAveSimNA ≤ 0.03) produced better-than-average overlay results for five of the seven template-dependent data sets (Figure 2). For a template, its δAveSimDY (or δAveSimNA) value equals the highest AveSimDY (or AveSimNA) among all of the templates in the same target data set subtracted by the AveSimDY (or AveSimNA) of the template. The average accuracy of alignment was improved from 20% to 35% for CDK2 and from 7% to 12% for HIV. Our template selection also minimizes the number of initial templates used to develop the final working 3D

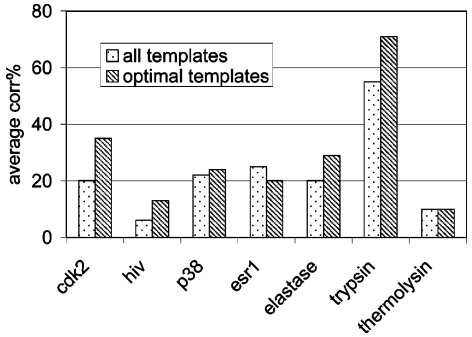

Figure 2. Average accuracy from ROCS−flexible for all templates and the optimal templates selected by the AveSim rule, i.e., δAveSimDY $\leq$ 0.06 and δAveSimNA $\leq$ 0.03, where AveSim refers to the average value of the similarity of all compounds to a template and δAveSim refers to the difference between such an average value and the largest AveSim found from all templates in the same protein data set. The number of optimal templates selected was 12 for CDK2, 5 for HIV, 6 for p38, 7 for ESR1, 4 for elastase, 3 for trypsin, and 4 for thermolysin.

pharmacophore hypothesis (e.g., 12 out of 57 for CDK2 or 5 out of 28 for HIV). This could complement the previously mentioned preselection protocols used by other programs such as Catalyst.[2,30]

3. Benchmark for Applicability and Predictions of Accuracy. All previous results utilized an assumption of knowledge of the X-ray bound conformation for templates. In a vast majority of drug discovery problems, we do not have the knowledge of an X-ray structure for any of the molecules. If such knowledge is available, docking[15] or similarity-driven docking[17] is usually the preferred method of generating the molecular overlays. In the following sections, we consider another remaining uncertainty of molecular overlays—the conformational deviation of the template from its X-ray complex. In this part of the study, we used 137 ligands from seven test systems: CDK2, HIV, ESR1, p38, thermolysin, elastase, and trypsin.

Flexibility Tolerance of Templates. To examine the allowable deviation of the template from its bioactive/X-ray conformation, we used 131 templates from seven test systems. Six templates from these systems were excluded because they did not produce any correct overlays even with their bioactive conformations. The results from flexible ROCS overlays with CORINA computationally generated template conformations (Figure 3) indicated that a 1.5 Å RMSD would give results not significantly different from those obtained with X-ray templates. The accuracy dropped significantly with a RMSD below 1.5 Å and was completely lost at 2.75 Å. Similar results were also obtained from flexible FlexS overlays and with the lowest-energy OMEGA conformations as templates (Table S3 and S4 of the Supporting Information).

Overlay Accuracy Prediction. We used all 137 compounds from seven test systems to evaluate the accuracy with no knowledge of X-ray structures of any molecules. We utilized computer-generated CORINA conformations for templates and ROCS flexible options for overlays. Our hypothesis was that the overall accuracy in this test would be related to the accuracy obtained with X-ray templates and the deviation of templates from the X-ray conformations. The hypothesis was consistent with the results as the average accuracy in this test was highly correlated ($r^2 = 0.96$) with the product

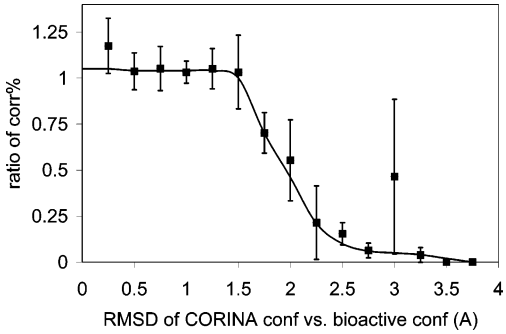

Figure 3. Effect of the conformational deviation of the template from its bioactive structure on the overlay accuracy for ROCS−flexible overlays. The corr% ratio of a template refers to the accuracy obtained from using its CORINA conformation as the template conformation divided by the accuracy from using its bioactive conformation. Templates were grouped into bins on the basis of the heavy atom RMSD of the CORINA conformation from the bioactive template conformation. A bin size of 0.25 Å was utilized. All templates with a RMSD >3.75 Å were included into the last bin. Each bin contains at least five templates, for a total of 131 templates. Every template gave at least one correct overlay when the bioactive conformation was used. The standard errors of the accuracy ratio are also shown. The outlier at the 3 Å bin is due to a high ratio of a single template 4hvp from HIV, whose accuracy jumped from 3.6% to 10.7%. The mean value and standard error would be 0.04 and 0.02, respectively, if 4hvp was not considered.

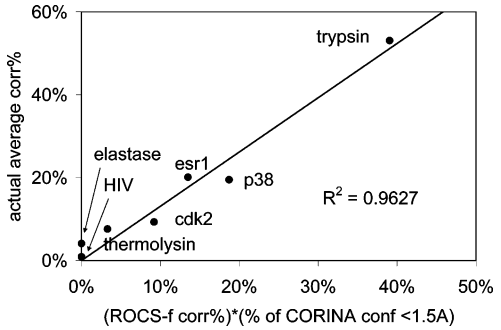

Figure 4. Actual alignment accuracy using CORINA conformations for the templates vs the product of the accuracy using the templates' bioactive conformation and the fraction of CORINA conformations within a 1.5 Å RMSD of the bioactive conformations. ROCS−f = ROCS−flexible.

of the fraction of templates with conformations close to those of X-ray results (below 1.5 Å) and the results with X-ray templates (Figure 4).

Our results suggested that real, unbiased overlay accuracy could be tackled independently with two independent solutions—one which focused on improving the conformational aspects of compounds and the second which improved on the overlay methods utilizing X-ray templates. In addition, our results indicate that, when the bioactive conformation of the template is not known, one might expect the overlay accuracy to be very low (<1%) for highly flexible compounds such as HIV ligands but reasonable for others, varying from 4% for elastase through 9% for CDK2 to 53% for trypsin. The results were similar to the observations made by Erickson et al.,[15] who compared the docking accuracy using the X-ray crystal structures from the ligand−protein complex with that from the apo protein and using the CORINA conformation for the ligands. The docking accuracy dropped from 56.7 and 18.5% to 36.7 and 0.9% for trypsin and HIV ligands, respectively.[15] Thus, a 3D overlay method appears comparable in accuracy to docking to unligated (apo)

structures but less accurate than docking to complexed structures.

CONCLUSIONS

The accuracy of molecular overlay depends on the choice of the template, the similarity of the molecules to the template, and the accuracy of the template's conformations. Both ROCS and FlexS could handle the flexibility of molecules reasonably well, provided the compounds were not highly flexible. We found that the accuracy of overlays could be separated into conformational aspects of the templates and the intrinsic accuracy of the overlay methods. We derived empirical rules to identify the compounds with a high and low probability of correct overlay accuracy. When a combination of various similarity measures was used as filters, some structurally diverse compounds could still be identified with a reasonable likelihood of being aligned correctly. This study also established a benchmark of expectations for molecular overlay accuracy which varies from 0 to 53% depending on the protein targets.

ACKNOWLEDGMENT

We thank Dr. Daniel Robertson for providing technical support, Dr. Jeffrey Sutherland for critical review and improvement of the readability of the paper, and the members of the Eli Lilly Global Computational Chemistry group for constructive comments and discussion.

Supporting Information Available: Data sets and overlay results. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES AND NOTES

(1) Kramer, A.; Horn, H. W.; Rice, J. E. Fast 3D molecular superposition and similarity search in databases of flexible molecules. *J. Comput.-Aided Mol. Des.* **2003**, *17*, 13−38.

(2) Sprague, P. W.; Hoffman, R. Catalyst pharmacophore models and their utility as queries for searching 3D databases. In *Computer-Assisted Lead Finding and Optimization*; Waterbeemd, H., Testa, B., Folkers, G. VHCA: Basel, Switzerland, 1990; pp 230−240.

(3) Korhonen, S.; Tuppurainen, K.; Laatikainen, R.; Perakyla, M. Comparing performance of FLUFF-BALL to SEAL-CoMFA with a large diverse estrogen data set: From relevant superpositions to solid predictions. *J. Chem. Inf. Model.* **2005**, *45*, 1874−1883.

(4) Patel, Y.; Gillet, V. J.; Bravi, G.; Leach, A. A comparison of the pharmacophore identification programs: Catalyst, DISCO and GASP. *J. Comput.-Aided Mol. Des.* **2002**, *16*, 673−681.

(5) Chen, H. F.; Dong, X. C.; Zen, B. S.; Gao, K.; Yuan, S. G.; Panaye, A.; Doucet, J. P.; Fan, B. T. Virtual screening and rational drug design method using structure generation system based on 3D-QSAR and docking. *SAR QSAR Environ. Res.* **2003**, *14*, 251−264.

(6) Chen, Q.; Wu, C.; Maxwell, D.; Krudy, G. A.; Dixon, R. A. F.; You, T. J. A 3D QSAR analysis of in vitro binding affinity and selectivity of 3-isoxazolylsulfonylaminothiophenes as endothelin receptor antagonists. *QSAR* **1999**, *18*, 124−133.

(7) Debnath, A. K. Generation of predictive pharmacophore models for CCR5 antagonists: Study with piperidine and piperazine-based compounds as a new class of HIV-1 entry inhibitors. *J. Med. Chem.* **2003**, *46*, 4501−4515.

(8) van Drie, J. H. Pharmacophore Discovery − Lessons Learned. *Curr. Pharm. Des.* **2003**, *9*, 1649−1664.

(9) Rush, T. S., III; 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.

(10) Lemmen, C.; Lengauer, T.; Klebe, G. FLEXS: A method for fast flexible ligand superposition. *J. Med. Chem.* **1998**, *41*, 4502−4520.

(11) *ROCS*; Openeye Scientific Software: Santa Fe, NM, 2005.

(12) Kearsley, S. K.; Smith, G. M. An alternative method for the alignment of molecular structures: Maximizing electrostatic and steric overlap. *Tetrahedron Comput. Methodol.* **1990**, *3*, 613−633.

(13) Mestres, J.; Rohrer, D. C.; Maggiora, G. M. MIMIC: A molecular-field matching program exploring applicability of molecular similarity approaches. *J. Comput. Chem.* **1997**, *18*, 934−954.

(14) Lemmen, C.; Lengauer, T. Computational Methods for the Structural Alignment of Molecules. *J. Comput.-Aided Mol. Des.* **2000**, *14*, 215−232.

(15) Erickson, J. A.; Jalaie, M.; Robertson, D. H.; Lewis, R. A.; Vieth, M. Lessons in molecular recognition: The effects of ligand and protein flexibility on molecular docking accuracy. *J. Med. Chem.* **2004**, *47*, 45−55.

(16) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235−242.

(17) Wu, G.; Vieth, M. SDOCKER: A method utilizing existing X-ray structures to improve docking accuracy. *J. Med. Chem.* **2004**, *47*, 3142−3148.

(18) *QUANTA;* Accelrys: San Diego, CA, 2006.

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

(20) *OMEGA*; Openeye Scientific Software: Santa Fe, NM, 2004.

(21) Bostrom, J.; Greenwood, J. R.; Gottfries, J. Assessing the performance of OMEGA with respect to retrieving bioactive conformations. *J. Mol. Graphics Modell.* **2003**, *21*, 449−462.

(22) Willet, P. Similarity-based approaches to virtual screening. *Biochem. Soc. Trans.* **2003**, *31*, 603−606.

(23) Klebe, G.; Abraham, U.; Mietzner, T. Molecular similarity indices in a comparative analysis (CoMSIA) of drug molecules to correlate and predict their biological activity. *J. Med. Chem.* **1994**, *37*, 4130−4146.

(24) *I. MACS Keys*; Daylight Chemical Information Systems: Aliso Viejo, CA, 2005.

(25) Vieth, M.; Siegel, M. G.; Higgs, R. E.; Watson, I. A.; Robertson, D. H.; Savin, K. A.; Durst, G. L.; Hipskind, P. A. Characteristic physical properties and structural fragments of marketed oral drugs. *J. Med. Chem.* **2004**, *47*, 224−232.

(26) Agresti, A. *Categorical Data Analysis*; Wiley: New York, 1990.

(27) JMP, version 4.0.4; SAS Institute Inc: Cary, NC, 1989−2001.

(28) Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning Theory*; Springer-Verlag: New York, 2001.

(29) Martin, Y. C.; Kofron, J. L.; Traphagen, L. M. Do structurally similar molecules have similar biological activity? *J. Med. Chem.* **2002**, *45*, 4350−4358.

(30) *Catalyst*, version 4.9; Accelrys Inc.: San Diego, CA, 2005.

CI060134H