J Comput Aided Mol Des (2008) 22:507–521
DOI 10.1007/s10822-008-9177-8

A ligand's-eye view of protein binding

Robert D. Clark

Received: 15 November 2007 / Accepted: 9 January 2008 / Published online: 24 January 2008

Abstract Docking tools created for structure-based design and virtual screening have also been used to automate ligand alignment for comparative molecular field analysis (CoMFA). Models based on such alignments have been compared with those obtained based solely on shared ligand substructures, but such comparisons have generally failed to distinguish between conformational specification (alignment in the internal coordinate space) and embedding in a shared external frame of reference (Cartesian alignment). Here, large sets of inhibitors were docked into two cyclooxygenase and two reverse transcriptase crystal structures, and the poses generated were evaluated in terms of the CoMFA models they produced. Realigning the conformers obtained by docking by rigid-body rotation and translation to overlay their common substructures improved model statistics and interpretability, provided the protein structure used for docking was reasonably appropriate to the ligands being considered.

Keywords CoMFA · COX-2 · Cyclooxygenase · HEPT · Ligand alignment · QSAR · Reverse transcriptase · Surflex-Dock · Bioactive conformation

Introduction

Comparative molecular field analysis (CoMFA) [1] and its sister technology, comparative molecular similarity analysis (CoMSIA[1] [2]), are arguably the most widely-used 3D QSAR methods available today. The two differ in the form of the molecular interaction fields used, but both involve correlating the distribution of field intensities across ligands with binding affinity or biological activity. The regression models produced allow researchers to predict the potency of other candidate structures. Perhaps more importantly, such models can also help chemists visualize where and how known structures might best be modified to enhance activity—or to reduce it, in cases where an undesirable activity such as metabolic lability or some side effect is involved.

To get a predictive CoMFA model for flexible molecules, one must first get them into comparable conformations and then get them into a common frame of reference. These are two separate forms of "alignment," one of which takes place in the internal coordinate space of each individual molecule and another which takes place in the Cartesian space shared by all molecules. Some closely related 3D QSAR methods side-step one or the other component of the "alignment problem." VolSurf [3], for example, is based on differences in field intensities *within* molecules, which avoids the need for a common Cartesian frame of reference. Topomer CoMFA, on the other hand, replaces alignment rules with fragmentation rules, with conformations within fragments set by applying generic torsion rules as one moves away from the fragmentation points [4].

Any "alignment free" method necessarily sets aside some potentially useful information. Careful alignment is

R. D. Clark (✉)
Tripos Informatics Research Center, 1699 South Hanley Road, Saint Louis, MO 63144, USA
e-mail: bclark@tripos.com

[1] For the systems considered here and other similar ones, CoMFA and CoMSIA generally yield quite similar results. In the interests of simplicity, only CoMFA results are described here, but similar conclusions should hold true for CoMSIA.

typically difficult and/or tedious to do, however, especially for large data sets, and it is rarely obvious how or when the extra effort is worth making.

In "classical" CoMFA, structures are aligned to a template molecule drawn from one of three sources: a bound ligand extracted from an X-ray crystal structure; some tightly bound ligand put into its lowest energy conformation; or the minimized structure of a conformationally constrained active compound (an extension of the active analog approach used in pharmacophore analysis [5]). Other structures in the data set can then be constructed "by hand." by sequentially modifying the template. Alternatively, the torsions about rotatable bonds in structures produced by 3D coordinate generation programs can be modified to match the torsions about analogous bonds in the template.

The initial conformations obtained are then generally relaxed to their respective local energetic minima, in part to allow for effects such as steric interactions in *ortho*-substituted aromatic rings. In addition, doing so is intuitively appealing, since it is reasonable to expect that ligands will bind in minimally strained conformations. Unfortunately, drug-like molecules often include bonds between sp^2- and sp^3-hybridized atoms that have rather broad torsional minima. Relaxation may introduce more noise than signal in such cases, particularly when it comes to localizing the effects of structural variation on activity. Then, too, recent studies have suggested that the amount of internal strain found in bound conformations is greater than had been assumed—or that our ability to assess it accurately is inadequate to the task [6–8].

Once final conformations have been determined, Cartesian alignment can be accomplished by rigid-body rotation and translation to minimize the sum-of-squares deviation between reference atoms in each ligand and the corresponding atoms in the template. Alternatively, pseudo-energy terms can be added to the applied force field to drive the ligand relaxation towards optimal structural overlap with the template. Such extra terms could be included as artificial "force constants" between corresponding atoms, but more often take the form of penalties for field mismatches, a technique known as "field fit" [1, 9, 10].

These procedures can be tedious and rather error-prone, especially when large data sets are involved. Moreover, they can be difficult to apply to structurally diverse data sets. Such considerations have prompted some researchers to explore more generalized and more readily automated methods. One is to use flexible 3D searching to fit each ligand to a pharmacophore hypothesis before running CoMFA [11–13]. The hypotheses used could in principle be based on complementarity to the structure of known binding sites, but this is not done very often; the hypotheses used for such pharmacophoric alignments are usually obtained from known ligands instead. When available protein structures are utilized for alignment, it is usually via docking [14–16] rather than by way of deduced pharmacophores. In a rather extreme case of cross-fertilization among 3D QSAR "worlds", the derived CoMFA model can then used as a refined scoring function for subsequent docking analyses [17].

Several permutations of the various alignment methods have been compared, with mixed results. In particular, the statistical properties of CoMFA models based on docking alignments have been compared with those based on substructural alignment [14–16]. But there is a hybrid approach which has not been adequately evaluated: using docking to obtain conformations, then using rigid-body substructural *realignment* to put those conformations into a common Cartesian coordinate system. Such an approach gives a ligand's-eye view of binding, in that the model obtained shows how the protein shifts—in position and in conformation—to accommodate changes in ligand structure. Figure 1 illustrates the difference in perspective obtained for three thrombin complexes. When the complexes are aligned by the α-carbon backbones of the protein (left), the position of the urazole core shared by all three ligands seems to vary with substitution. When the complexes are instead aligned instead so as to optimize the overlap of the shared ligand core (right), it becomes

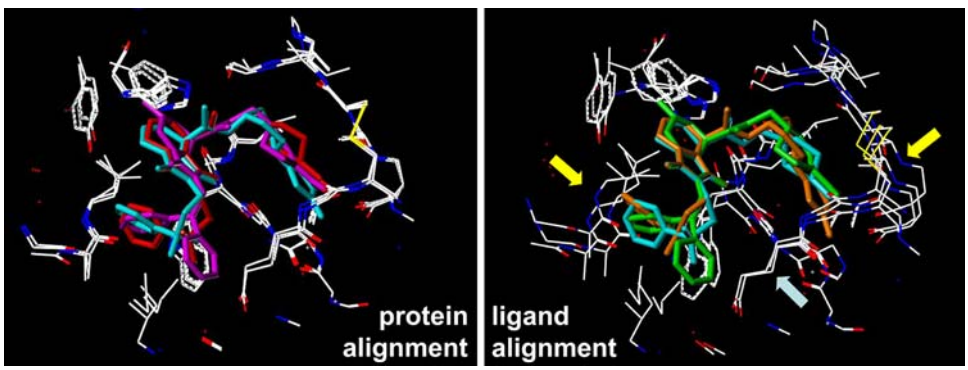

Fig. 1 Active sites in thrombin complexes 1C4W, 1D6W and 1D9I aligned by protein α-carbons (*left*) or by the urazole core common to all ligands (*right*)

apparent that some parts of the protein (e.g., those highlighted by the yellow arrows) shift substantially, whereas other parts of the protein (e.g., indicated by the cyan arrow) look the same from both points of view.

This paper compares and contrasts the protein- and the ligand-based perspective in terms of the CoMFA models generated from the corresponding alignments, and compares both to those for models derived from "pure" substructural alignment—i.e., a priori alignment.

Data sets

Cyclooxygenase

The inducible form of cyclooxygenase (COX-2) has been a major focus of drug discovery and development efforts almost since it was first identified as being a major player in inflammatory response. It is structurally distinct from the constitutive form (COX-1) involved in the gastrointestinal irritation caused by older non-steroidal anti-inflammatory drugs (NSAIDs) [18]. Several pharmaceutical companies set out to establish distinct patent estates, which led to a wide range of inhibitor classes being synthesized and several different crystal structures being obtained. This in turn produced a wealth of QSAR literature involving CoMFA [15, 16, 18–21] as well as a other descriptors and regression techniques [22–27].

The work described here was carried out on 304 COX-2 inhibitors from a list originally compiled by Chavatte et al. [28]; a trifluorohydroxypropyl pyrrole included there[2] has been omitted here because it represents a mixture of enantiomers whose individual activities have not been characterized. Compound **1** served as the conformational template for the "a priori" COX-2 alignments, with both pendant rings placed perpendicular to the central ring and the most-substituted side of **C** ring coming up out of the page in the orientation shown in Fig. 2 [29].

The representative structures shown in Fig. 2 span the five basic structural classes included in this data set. The highlighted bonds indicate the atoms used for alignment within and between classes and the circled atoms are identified with the **A** ring. The rough symmetry of the molecules necessitated a further level of orientational specification, however. Based on previous cross-class regression analysis [29], the default (*recto*) orientation was presumed to be the one that overlaid the respective phenylsulfonyl groups, regardless of how the central (**B**) ring was oriented. When neither pendant ring bore a sulfonyl group (e.g., **9a**–**c**) or both did (**5**), the **A** ring was assigned by matching to the core structure in the class template. All Cartesian alignments for the a priori COX-2 model were assigned in this way except for compound **9c**. There the distal methoxy group was expected to interact in much the same way the sulfonyl group in **10a** does rather than behaving like the *p*-methoxy group in **10b**. Therefore the alternative flipped (*retro*) orientation was used in the a priori alignment for **9c**.

A typographical error in published supplementary material [28] led to the mistaken use of a partially reduced structure for DuP-697 in our earlier work on this data set; the correct structure (Fig. 2) has been used here. Because the data set is so large, correcting this single structure had no appreciable effect on the cross-validation (CV) behavior of the CoMFA models derived from those alignments [29, 30].

The 3D structures corresponding to the a priori COX-2 alignment are available from the author on request.

Reverse transcriptase

The discovery that zidovudine (AZT) inhibits the reverse transcriptase (RT) of the human immunodeficiency virus type 1 (HIV-1) opened the door to effective treatment for acquired immune deficiency syndrome (AIDS). The next major advance in the area came with the discovery of the non-nucleoside reverse transcriptase inhibitors (NNRTIs) [31, 32]. In particular, the discovery that nevirapine (Fig. 3) bound at an allosteric site distinct from the catalytic region of the viral enzyme [33] quickly led to the realization that the "acyclouridine" inhibitors typified by 1-[(2-hydroxyethoxy)methyl]-6-(phenylthio)thymine (HEPT; Fig. 3) [34, 35] had the same mode of action [36, 37].

Publication of a large number of well-characterized analog structures along with biological activities measured under consistent conditions and the availability of public crystal structures have made HIV-1 RT a prime target for QSAR studies. Several CoMFA studies have been published [38–40] as well as a range of other analytical approaches [41, 42].

Luco and Ferretti considered 80 active analogs along with 27 inactive ones, using connectivity indices to build a classification model that was tested against 15 other compounds. Kireev et al., on the other hand, looked to descriptors based on properties dependent on 3D structure, but not to CoMFA. They performed extensive conformational analysis on 87 analogs, using dipole moments and associated 3D geometrical descriptors to create a set of linear regression models [43]. One of their primary conclusions was that energetically accessible HEPT conformations fell into two major families that could be characterized as being 1,6-*cis* or 1,6-*trans*, depending on the relative disposition of

[2] Their compound **18**.

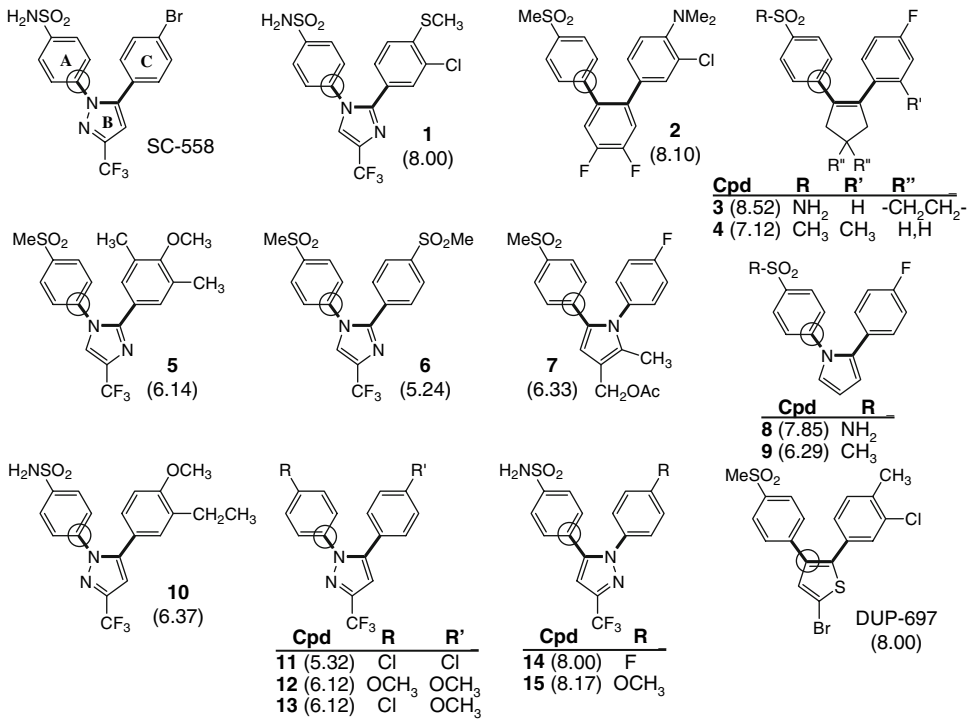

Fig. 2 Class templates and other representative structures from the 304 included in the COX-2 data set. Bold bonds represent the substructures used for a priori alignment and for realignment of docked poses. The circled atoms were overlaid in the canonical *recto* orientation

their *N* and *C* substituents with respect to the central ring. They concluded that 1,6-*trans* was the more likely binding mode, a prediction that was borne out experimentally [37, 44].

Clark et al. [13] merged the data set compiled by Kireev et al. [43] with that of Luco and Ferretti [45] to produce a set of 103 structures, including both axial and equatorial conformations of **12**. CoMFA models derived from a priori alignments of these compounds were used to validate the progressive scrambling approach to model validation [13]. The a priori conformations used for these compounds were obtained by modifying the torsions in the MKC-442 configuration produced by CONCORD [46, 47] to match the 1,6-*trans* configuration found in the 1RT1 crystal structure [37]. This template was then "morphed" into other analogs by incremental structural modification, with larger substituents "grown out" in maximally extended conformations [6]. Cartesian alignment was based on the pyrimidine core structure.

Several representative analogs are shown in Fig. 3. Full 3D structures for all 103 compounds in the HEPT data set are available from the author on request.

Methods

The initial 3D conformations were generated for all ligands from the respective 2D structures using CONCORD. For a priori alignments, systematic torsional modifications were made using scripts written in the SYBYL Programming Language (SPL), with the relevant substructures specified in SYBYL Line Notation (SLN [48]) as queries. Secondary templates for each substructural class were aligned to each data set's master template using the Fit Atoms command, then ligands in each class were aligned to the respective class template using the Database Align command in SYBYL [49]. Partial atomic charges were assigned using methodology described by Gasteiger and Marsili [50] extended to include π electrons as well as the σ bonding framework ("Gasteiger–Hückel charges").

Structures for protein–ligand complexes with PDB accession codes 1CX2 and 6COX [18] were downloaded via the Protein Data Bank website maintained by the RCSB[3] [51], as were those for 1VRT [36] and 1RT1 [37]. Extraneous protein chains were removed, as were all waters not located in or near the binding site. Subsequent protein target preparation was carried out using the generalized docking interface in a pre-release version of SYBYL 8.0. This included adding hydrogens to each complex, extracting bound ligands and generating a protomol for use in the subsequent Surflex-Dock runs [52, 53]. Default parameters were used for generating protomols and for docking itself, except that the number of restarts was set to three and the number of configurations considered for each fragment was increased from 20 to 50. Ring flexing

[3] The Research Collaboratory for Structural Bioinformatics (http://www.rcsb.org/pdb).

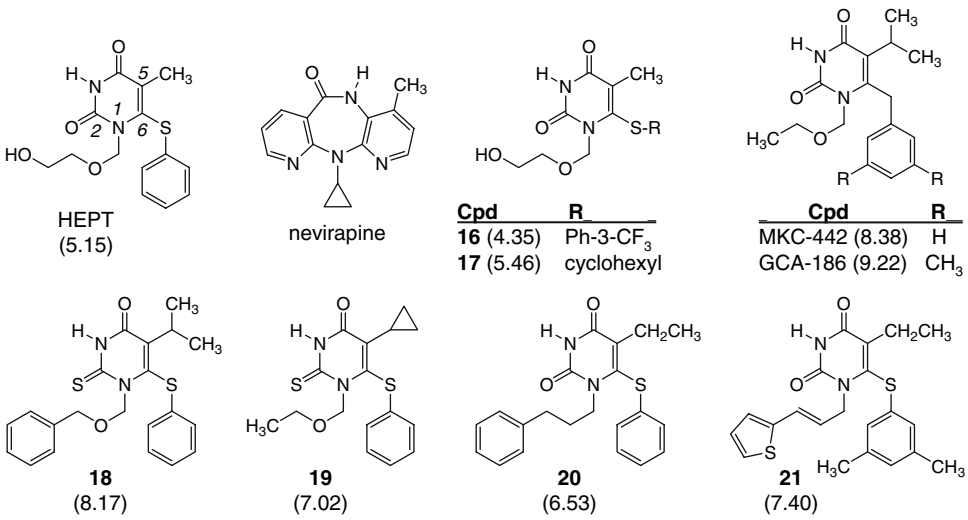

Fig. 3 Representative inhibitors taken from among the 103 in the RT data set

was turned on, as were ligand pre- and post-optimization [54].

The ten highest-scoring poses were retained. CoMFA models for docked alignments utilized the target protein as the common frame of reference, and the highest-scoring pose was used in all cases to determine the Surflex-Dock alignments.

The same procedure was carried out to rigidly realign docked poses as to do Cartesian alignment a priori. The coordinate origin for the frame of reference was established by centering the master template molecule in its docked conformation and establishing X, Y and Z coordinate axes along the master template's principal axes, then fitting specified atoms in other docked ligand poses to that template. All HEPT analogs were realigned to the six atoms in the pyrimidine core of GCA-186 (Fig. 3). COX-2 inhibitors were realigned to class templates using the substructures highlighted in Fig. 2, with **4** serving as the master template. The sense of the mapping used (i.e., which pendant ring's root atom was to be identified with circled anchor in the template) was based on the coarse alignment determined by docking. The "*recto*" alignments were those that directed the nominal **A** ring into the α pocket of the protein and the **C** ring into the γ pocket. The "*retro*" configurations, in contrast, had the **B** ring flipped so as to position their **A** ring in the γ pocket and their **C** ring in the α pocket. Ligands docking in *retro* configurations were realigned by mapping the base atom in their **C** ring onto the **A** ring of the template and the base atom of their **A** ring onto the **C** ring of the template.

In some cases, the alignment of the highest-scoring pose diverged substantially from that of the corresponding template. In that event, the other nine high-scoring ligand poses were examined and the best pose involving one of the "orthodox" (*recto* or *retro*) orientation was subjected to re-alignment instead. If all of the poses deviated from the canonical alignments, the offending ligands were omitted from the realigned models.

CoMFA fields were calculated using default settings unless otherwise indicated, with steric and electrostatic interaction energies calculated at each lattice point with respect to an sp^3 carbon probe atom bearing a net charge of +1. Two rectilinear region files with lattices extending 4 Å beyond the edge of any structure in the data set were created for each alignment. One analysis was done using a region file that defined a standard (cubic) type of lattice with 2 Å spacing along each coordinate. A second analysis was done using fields calculated on an orthogonal, face-centered cubic lattice constructed with 2.5 Å spacing.[4] Comparing the results for models constructed using such complementary lattices provides a good test for model dependence on placement and orientation that avoids potential complication by edge effects that may be encountered when ligands are rotated or translated on a fixed lattice [55].

In PLS, one can only be confident that one model is statistically superior to another if its standard error (SE) is lower and its q^2 is higher *and* the first models model is at least as simple (has as few or fewer components) than does the second. For this reason, cross-validation results are reported at three different levels of complexity for each model. Three-fold cross-validation entails partition of the data set into three subsets, each of which then serves as a test set. The reported standard errors of cross-validation (SE_{CV}) are root mean squares across four independent runs, from which the reported q^2 values were calculated.

[4] The FCC lattice is generated by nesting one cubic lattice inside another, offset by a half step along each axis. The TAILOR AUTO_REGION LATTICE and TAILOR AUTO_REGION STEP_SIZE commands controls the relevant settings in SYBYL.

Progressive scrambling [13, 30] was carried out in using default settings: responses were partitioned into 2–10 bins, and responses within each bin were randomized 10 times for each partition. A critical perturbation point of 0.85 (i.e., 15% added noise) was used. As for SE_{CV}, the corrected standard errors of prediction (cSDEP) are reported as the root mean squares from replicate runs—two, in this case. The extrapolated predictive correlation coefficient (Q_0^2) cannot be calculated directly from cSDEP and so was simply averaged across the duplicate runs.

Results

The individual inhibitor potencies for these target proteins were not generally well-differentiated in terms of Surflex-Dock score, but the predicted potencies were reasonably accurate for three of the four systems considered here. For the COX-2 data set, the experimentally observed mean pIC50 was 7.12 ± 0.93 (mean ± SD) vs Surflex-Dock scores of 7.58 ± 0.86 and 7.88 ± 0.87 for docking into apo1CX2 and apo6COX, respectively. The similarity in spread between the observed and predicted values suggests that differences in affinity for these ligands fall below the resolution of the method. Higher accuracy was not in fact anticipated, given the relatively small variation in potencies of the ligands (less than one log unit in standard deviation); the use of IC50s in lieu of K_i values; and the known plasticity of the binding sites [56, 57].

The corresponding predicted affinity values for the HIV-1 RT inhibitors examined were 6.50 ± 1.32 (observed) vs 5.25 ± 3.06 (1VRT) and 7.46 ± 2.01 (1RT1). The exaggerated range for docking into apo1VRT results from large Surflex-Dock clash scores for these ligands. Indeed, one ligand was predicted to have an affinity of 2.4×10^9 M, a deviation consistent with literature indications that a rigid protein model may not be realistic in this case [37, 43]. The range seen for 1RT1 is much more consistent with the experimental range, though some excessive clash is still evident. For the studies described here, however, the relative positioning of the different ligands was of more interest than scoring. Moreover, the problem would not necessarily be obvious in a "real-world", prospective application. Hence both alignments were carried forward to see if it could be detected after the fact.

The majority of the binding modes returned by Surflex-Dock were consistent with the crystal structures available for the corresponding template ligands. As is discussed in detail below, the best-scoring pose for a few of the 304 COX-2 ligands deviated markedly from those of the templates. These were all structural outliers which may well exhibit a different binding mode [56, 57].

Docking into reverse transcriptase

The apo1VRT and apo1RT1 binding sites are quite different (Fig. 4). The enzyme tends to collapse around the ligand in the allosteric binding site, and the ligand in 1VRT (nevirapine) is more compact than many HEPT analogs—e.g., **13**, **15** and **16** (Fig. 3). The ligand in 1RT1 is MKC-442, which is larger than nevirapine and binds in a relatively extended conformation. Several residues around the binding site rearrange themselves as a result. In particular, shifts in the position of Lys102 and Lys103 open a pocket in the "floor" of the binding site, a pocket that can receive the N^1-substituents and allow HEPT analogs to take on their characteristic [43] 1,6-*trans* binding conformation. Surflex-Dock is unable to fit the ligands comfortably into the more constrained apo1VRT binding site, however, and many end up in a 1,6-*cis* conformation or an alternative 1,6-*trans* configuration in which the N^1-substitutent comes "upwards" from the perspective of Fig. 4, rather than being directed "downward."

In fact, the ligands are so tightly packed into the 1VRT site that no clear substructural alignment is apparent in the docked structures at all. The generality of their *cis* configuration only becomes evident after the docked ligands have been realigned based on the pyrimidine substructure they share (Fig. 5). It is not clear from the 1RT1 results whether the binding site might be able to expand even further so as to comfortably accommodate the fully extended conformation of compounds such as thiophene **21** presumed in the a priori alignment.

Realignment of the docked conformations yields much more crisp overlays than does docking itself, and the a priori alignment is crisper still. This point is best appreciated when one realizes that each set of structures in Figs. 4 and 5 is comprised of the same ten ligands.

Figure 6 shows the field contours for models based on the five different alignments considered here. Yellow isopleths indicate regions with relatively large negative steric regression weights[5], i.e., regions in which substitution is associated with reduced affinity. Regions for which the steric regression weights were relatively large and positive—i.e., where added bulk is associated with an increase in affinity—are colored gray. Negative electrostatic contours are shown in red and blue indicates regions where electrostatic regression weights are relatively large and positive. In other words, red isopleths show where negative charge is favored over positive charge, whereas blue contours indicate where positive charge "favors" binding.

[5] Here, "regression weight" is defined as the product of the regression coefficient and the standard deviation of the field at each lattice point.

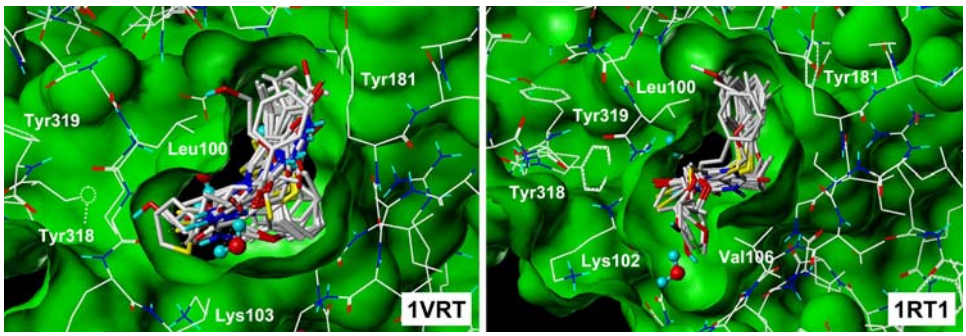

Fig. 4 Representative poses obtained using Surflex-Dock to dock HEPT analogs into apoenzyme from 1VRT (*left*) or 1RT1 (*right*)

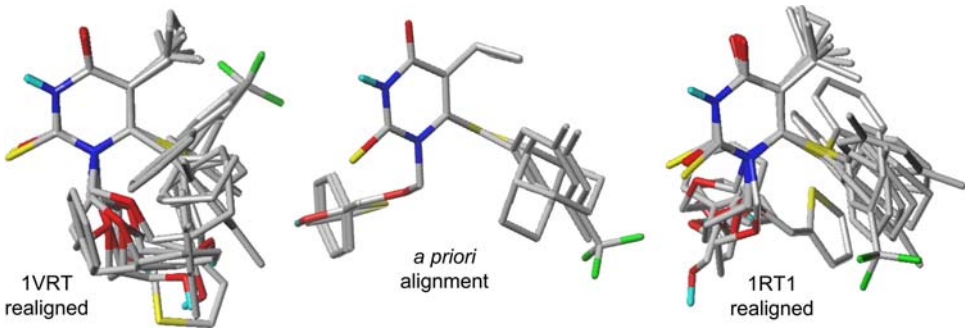

Fig. 5 Docked poses from 1VRT (*left*) and 1RT1 (*right*) realigned by their pyrimidine cores. The a priori alignment (*center*) is included for comparison

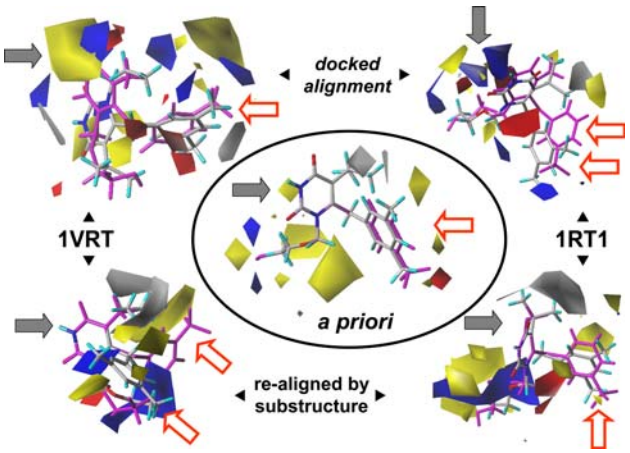

Fig. 6 Steric and electrostatic CoMFA contours plotted at the 20th and 80th percentile levels for docking poses (*top*) and substructural realignments (*bottom*). Results based on 1VRT are shown at left and those based on 1RT1 are shown at the right. The results for the a priori alignment (*center*) are included for comparison. Gray contours indicate regions where increased steric bulk is correlated with increased activity, whereas yellow contours indicate areas in which steric bulk is disfavored. Red contours highlight regions in which more negative charge is favored, and blue contours highlight those in which more positive charge is associated with high activity

In general, the realigned models focus on the structural variability in substituents at N^1, C^5 and C^6. The corresponding docked poses, in contrast, have major contours associated with the pyrimidine ring, which varies only in its C^2 (thio)carbonyl functionality. Indeed, the realigned models are superior to the a priori alignment in this regard. Though isopleths in this area (indicated for each model by a solid gray arrow in Fig. 6) accurately reflect the algebraic reality of the SAR, they do not provide useful feedback to the chemist.

The model contours in Fig. 6 are superimposed on the structures of GCA-186 and of compound **16**. With a pIC50 of 9.22, the former is the most potent ligand in the data set. The latter, which is highlighted in magenta, is among the weakest binders (pIC50 = 4.35). It bears a 3′-trifluoromethyl group and its lack of potency is typical of compounds that bear a 3′-substituent but no 5′-substuituent. This makes the contrast between it and the 3′,5′-dimethyl GCA-186 particularly instructive. The phenyl rings (open red arrow) are overlaid in the a priori alignment, and the model contours in that region provide no rationale for the observed difference in activity. They are, in fact, contrary: there is a negative steric (yellow) region near the 5′-methyl group and a negative electrostatic contour near the 3′-position which should favor the trifluoromethyl group in **16**. In this model, the observed difference in activity is accounted for instead by differences at other variation sites—in particular, whether or not the N^1-substituent bears a terminal hydroxyl group and the size of the C^5 substituent.

The rings are also superimposed in the docked 1VRT alignment, which yields a complex set of contours in this area that is very hard to interpret. The docked alignment for 1RT1 and the realigned 1VRT model position the rings very differently for each ligand. Probably the most useful model is the one derived from the substructurally realigned 1RT1 conformations, in which the rings are overlaid but

rotated such that the 3′-CF3 group of **16** projects into two negative steric regions (Fig. 6).

Docking into cyclooxygenase

The differences in the 1CX2 and 6COX binding sites are more subtle than are those between 1VRT and 1RT1, in part because the bound ligands are the same in both cyclooxygenase complexes. The differences are real, however, as can be seen by comparing the best-scoring poses obtained for 16 ligands selected to span the range of structure and affinity in the data set (Fig. 7).

The binding motifs predicted by Surflex-Dock fall into the same two main clusters for both targets, but the detailed interactions differ. Unsubstituted cyclopentenes such as **3b** sit "higher" in the β pocket of the binding site than do most of the analogs. The **A** rings (phenylsulfonyl groups) penetrate more deeply into the α pocket near Ser353 in the latter case than in the former, less common one. The shift is more pronounced for 6COX than for 1CX2, but is clear for both targets. Key hydrogen bonds are formed in the α pocket between the critical sulfonyl and sulfonamide oxygen atoms and the backbone amide of Phe518, and between the sulfonamide NH and the backbone carbonyl of Arg513. Hydrogen bonds to Ser353 and His90 are also evident, with these amphiphilic residues acting as acceptors in the complexes with 6COX and as donors to the sulfonyl oxygens in 1CX2 (Fig. 4).

In most cases, the sulfonyl "warhead" of each bound ligand was positioned in the α pocket in the best-ranked pose, which were the poses used in the docked alignments. This was not true in all cases, however. Figure 8 shows the structures of fifteen ligands for which the highest-scoring pose did not match the expected orientation when docking into apo1CX2. In some cases (e.g., **26**), all of the ten best poses were *retro*. Good poses for others (e.g., **27** and **35**) were a mix of *retro* and *recto* orientations. For several inhibitors, the entire ligand was rotated, with the **B** ring placed in the α (**22**) or γ (**23**, **28**, **32** and **33**) pocket, a binding mode not inconsistent with data obtained for other ligand classes [56, 57]. In the most extreme cases (**29**, **30** and **36**), the central **B** rings were rotated around the bond to the **A** ring by about 100°, with the result that that the **C** ring projected "back" into the binding site and the γ pocket went unoccupied.

Three COX-2 inhibitors—compounds **11**, **12** and **13**—lack phenylsulfonyl groups. Compound **13** docked preferentially in a *retro* configuration, with the methoxy group taking the place in the α pocket "usually" occupied by the sulfonylgroup—just as had been anticipated in the a priori alignment (see above). What was not expected was that the pyrazole cores in **11** and **12** would also be flipped into the *retro* orientation by Surflex-Dock, as was the imidazole core in all high-scoring poses for the *bis*-phenylsulfonyl **6**. In both cases, the *C*-phenyl (**C**) group—not the *N*-phenyl (**A**) group expected a priori—was positioned in the α pocket. In none of these cases would the docked orientation have been anticipated based on consideration of substructure alone.

The *retro* or *recto* pose with the best Surflex-Dock score was used for realignment; which was used for each deviating ligand is indicated in Fig. 8. All poses for **30** had essentially the same unorthodox orientation, so that ligand was omitted from the realignment analysis for 1CX2. Hence $N = 303$ for those analyses.

Comparable results were obtained for docking into apo6COX, though the particular ligands that deviated differed somewhat. The *des*-sulfonyl inhibitor **12** was predicted to bind in a *retro* orientation, just as for 1CX2. All good poses for the **11** and **13** had the ligand rotated so as to place the central pyrazole ring in the α pocket rather than the β pocket, however. The same binding mode accounted for all high-scoring poses of **30** and **38**. The best pose for the other 4,4-dimethylcyclopentene derivative in the data set—compound **37**—was rotated in the same way, so the second best (*recto*) pose was used for realignment. As indicated in Fig. 9, the ligands with unorthodox docking results yielded predominantly *retro* orientations. All told, four ligands (**11**, **13**, **30** and **38**) yielded no poses suitable for realignment and so were dropped, leaving a ligand count of $N = 300$ for the subsequent 6COX CoMFA.

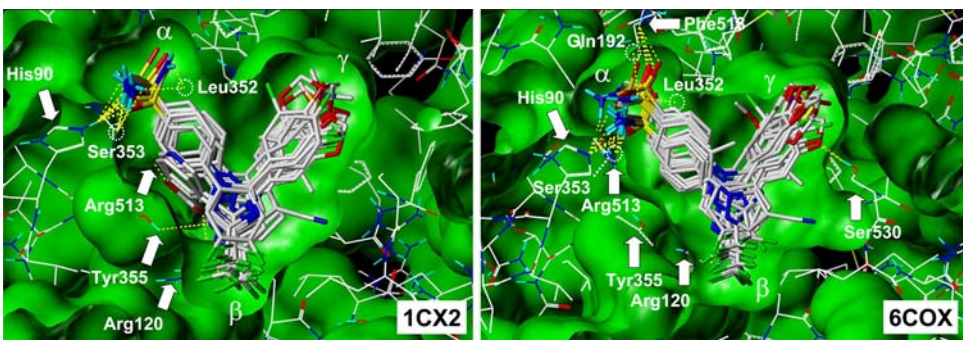

Fig. 7 Representative poses for COX-2 inhibitors docked into the apoenzymes from 1CX2 (*left*) and 6COX (*right*)

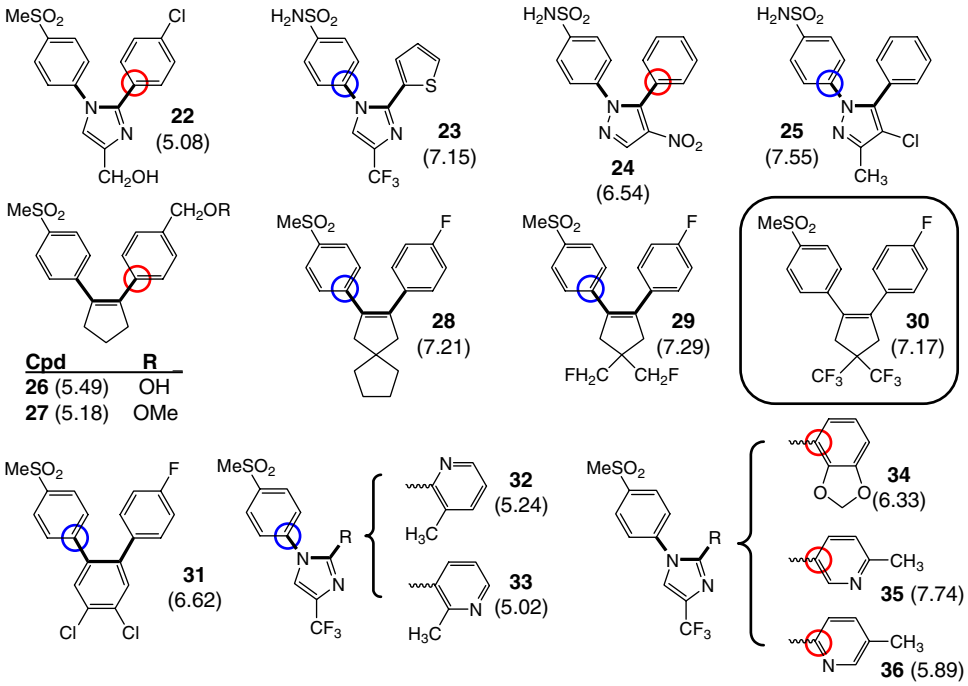

Fig. 8 COX-2 inhibitors for which the ligand positioning in the best-scoring pose for 1CX2 was contrary to that expected. Circled atoms were overlaid with those at the base of the **A** ring in the corresponding class template (Fig. 2). Blue indicates that the best-scoring orthodox pose exhibited *recto* orientations and red indicates that the best-scoring orthodox pose was *retro* in orientation

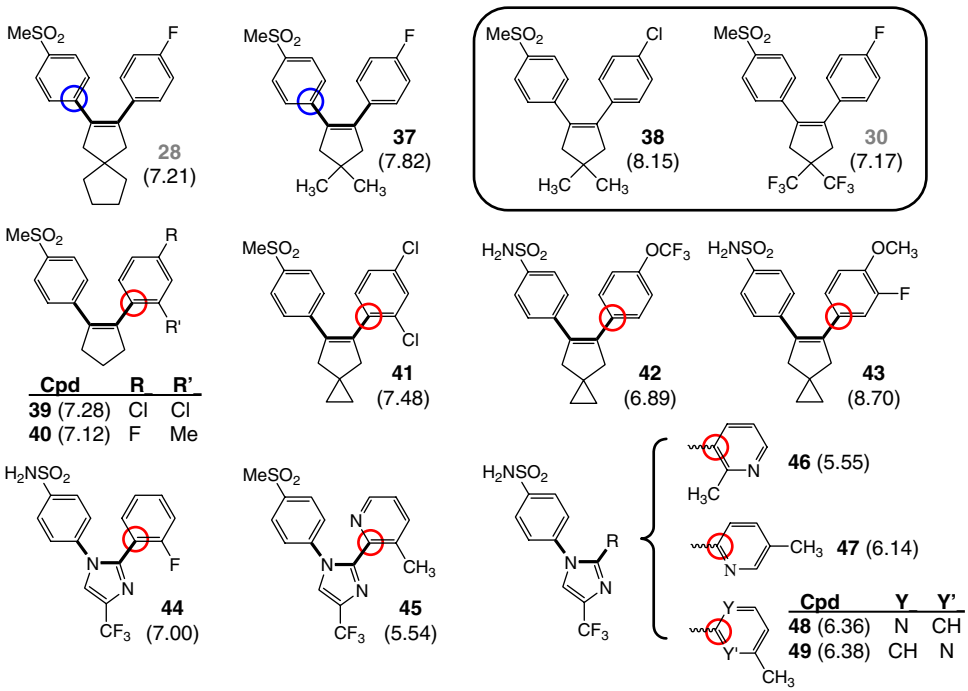

Fig. 9 COX-2 inhibitors for which the ligand positioning in the best-scoring pose for 6COX was contrary to that expected. Circled atoms were overlaid with those at the base of the A ring in the corresponding class template. Blue indicates that the best-scoring orthodox pose exhibited *recto* orientations and red indicates that the best-scoring orthodox pose was *retro* in orientation

The COX-2 inhibitors in each class that had high CScore [58, 59] and Surflex-Dock ranks were used as class templates; these are the analogs shown in Fig. 2. Figure 10 shows the two sets of realigned docked conformations for the same 16 representative inhibitors shown in Fig. 7, along with the corresponding a priori alignment. Though the various torsional and translational displacements seen in the docked alignments may appear random, they are in fact highly correlated. As a result, much of the configurational "noise" disappears when the docked conformations are realigned based on substructure. The overlays produced by realigning the docked conformations by their core substructures—i.e., by imposing a ligand-based perspective—produce much crisper alignments than do those in which the protein is fixed in place. Nonetheless, they are obviously less sharply focused than is the a priori alignment shown in the center of Fig. 10.

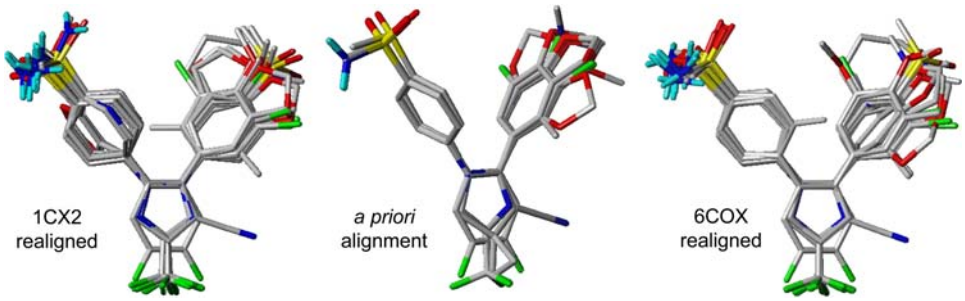

Fig. 10 Docked poses from 1CX2 (*left*) and 6COX (*right*) realigned by substructure. The a priori alignment (*center*) is included for comparison

Figure 11 shows the interaction contours from the CoMFA models derived from the five COX-2 alignments considered here. In qualitative terms, the model derived directly from the docked poses obtained using apo6COX (top left) is the most coherent; that obtained by docking into apo1CX2 (top right) and the realigned 6COX model (bottom left) and are similar in complexity to each other but more tightly focussed than is the 6COX model; the realigned 1CX2 model (lower right) is simpler still; and the a priori alignment (center) is simplest of all. Note the progressive simplification around the parts of the inhibitors highlighted with gray and stippled arrows directed towards the **A** and **B** rings, respectively, where the inhibitors vary least in structure. Note, too, the increasing focus around the bottom of the **B** ring (open red arrow), from which substituents project down into the β pocket. Structural variation there is a major determinant of affinity.

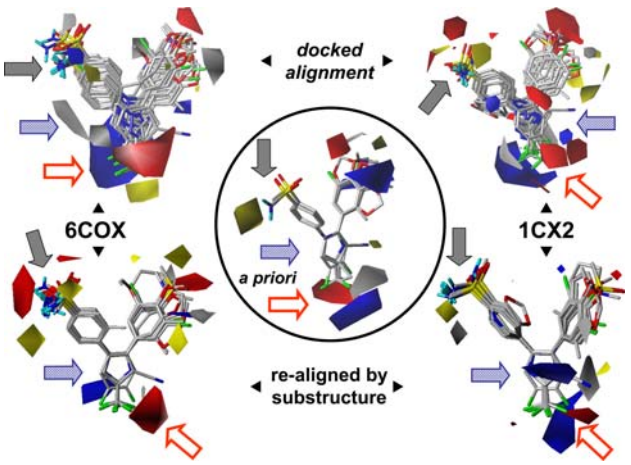

Fig. 11 Steric and electrostatic field contours for CoMFA models based on docking poses (*top*) and substructural realignments (*bottom*). Results based on 6COX are shown at left and those based on 1CX2 are shown at right. The results for the a priori alignment (*center*) are included for comparison. Gray contours indicate regions where increased steric bulk is correlated with increased activity, whereas yellow contours indicate areas in which steric bulk is disfavored. Red contours highlight regions in which more negative charge is favored, and blue contours highlight those in which more positive charge is associated with high activity. The contours shown were plotted at the 15th and 85th percentile levels for clarity; analogous plots at the default levels of 20 and 80% were too visually cluttered to be useful for the 6COX docked alignment

Regression statistics

Some of the fuzziness in the docked and realigned overlays reflects real interactions with protein residues and some can be attributed to essentially random inaccuracies in the crystal structure or approximations made by Surflex-Dock. One must then ask: how much of the variation is information and how much is noise? One way to address this question is to use the CoMFA model statistics to assess the degree of internal consistency in the alignments.

The PLS statistics set out in Table 1 provide quantitative support for the qualitative conclusions described above. The a priori alignment yields the lowest cross-validated standard error (SE_{CV}) and the highest q^2, whether evaluated by predicting each individual activity using a model based on the balance of the data set ("leave-one-out" or LOO crossvalidation), or whether one-third of the observations are held back as a test set and predicted using a model based on the other two-thirds of the observations ("three-fold" crossvalidation). Overall, the simple CV statistics for the COX-2 data set yield a rank-ordering of: a priori > realigned 1CX2 > 1CX2, realigned 6COX > 6COX.

PLS regression statistics obtained without crossvalidation are a measure of goodness-of-fit, which includes systematic bias such as deviations from linearity as well as random error. Here the rank-order is somewhat different, indicating that the underlying assumption of linearity holds up at least as well for the realigned conformations obtained from docking as for the a priori alignment: realigned 1CX2 > a priori, realigned 6COX > 1CX2 > 6COX.

Redundancy in a data set can seriously inflate the LOO q^2 of a model, especially in data sets as large as those considered here [13]. Three-fold crossvalidation is a more rigorous test of predictivity than is LOO, but the effective data set size is then reduced to 2/3*N*. The predictive power of such reduced models is necessarily compromised to some degree, leaving one at risk of seriously underestimating the predictivity of a full, uncrossvalidated model

Table 1 PLS statistics for CoMFA models. Values highlighted in bold indicate the best model by each criterion

Data set	COX-2[a]						HIV-1 RT[b]					
Conformation	A priori		1CX2		6COX		A priori		1VRT		1RT1	
Alignment	Substr.[c]		Surflex[d]	Substr.	Surflex	Substr.[e]	Substr.		Surflex	Substr.	Surflex	Substr.
Leave-one-out cross-validation (LOO)												
SE_{CV}	4[f]	0.621	0.707	0.641	0.735	0.705	3[g]	0.569	0.979	0.971	**1.102**	0.919
	5	0.601	**0.687**	0.637	0.731	0.704	4	**0.563**	0.938	**0.974**	1.144	**0.931**
	6	**0.594**	0.689	**0.627**	**0.728**	**0.702**	5	0.585	**0.933**	1.014	1.147	0.941
q^2	4	0.562	0.432	0.539	0.386	0.436	3	0.819	0.464	0.473	**0.321**	0.528
	5	0.591	**0.465**	0.546	0.396	0.440	4	**0.824**	0.513	**0.475**	0.277	**0.536**
	6	**0.602**	0.465	**0.562**	**0.402**	**0.445**	5	0.812	**0.524**	0.437	0.280	0.526
Three-fold cross-validation												
SE_{CV}	4	0.639	0.721	0.666	**0.759**	**0.725**	3	0.635	0.932	1.130	**1.195**	0.961
	5	0.616	**0.711**	**0.659**	0.765	0.727	4	0.612	**0.906**	**1.112**	1.232	**0.956**
	6	**0.607**	0.723	0.665	0.784	0.739	5	**0.606**	0.927	1.144	1.221	0.977
q^2	4	0.536	0.410	0.502	**0.345**	**0.404**	3	0.775	0.515	0.287	**0.202**	0.484
	5	0.570	**0.428**	**0.515**	0.337	0.402	4	0.793	**0.546**	**0.316**	0.160	**0.494**
	6	**0.584**	0.410	0.508	0.307	0.386	5	**0.799**	0.530	0.284	0.183	0.477
Without cross-validation												
SE	4	0.545	0.557	0.556	0.621	0.576	3	0.507	0.530	0.470	**0.574**	0.503
	5	0.519	**0.517**	0.513	0.571	0.543	4	**0.444**	0.470	**0.406**	0.493	**0.455**
	6	**0.500**	0.492	**0.476**	**0.541**	**0.504**	5	0.403	**0.393**	0.332	0.415	0.399
r^2	4	0.663	0.648	0.653	0.562	0.623	3	0.856	0.844	0.878	**0.818**	0.860
	5	0.695	**0.698**	0.705	0.631	0.666	4	**0.891**	0.879	**0.910**	0.867	**0.886**
	6	**0.717**	0.727	**0.748**	**0.669**	**0.713**	5	0.911	**0.916**	0.940	0.907	0.914
Progressive scrambling												
cSDEP[l]	4	0.681	0.743	0.698	0.779	**0.739**	3	**0.741**	1.054	**1.036**	**1.136**	**1.015**
	5	0.670	**0.732**	0.688	**0.778**	0.743	4	0.758	**1.050**	1.040	1.179	1.015
	6	**0.669**	0.737	**0.682**	0.788	0.749	5	0.801	1.058	1.084	1.191	1.035
Q_0^{2j}	4	0.556	0.439	0.533	0.365	**0.447**	3	**0.814**	0.445	0.470	**0.327**	**0.529**
	5	0.578	**0.464**	0.553	**0.370**	0.441	4	0.800	**0.458**	**0.472**	0.269	0.503
	6	**0.582**	0.455	**0.566**	0.352	0.433	5	0.758	0.455	0.418	0.261	0.483
$\frac{dq^2}{dr^2_{yy'}}$k	4	0.701	0.360	0.528	0.559	0.533	3	**0.949**	0.507	0.455	**0.491**	**0.839**
	5	0.735	**0.512**	0.580	0.594	0.580	4	1.085	**0.966**	0.617	0.241	1.074
	6	**0.814**	0.475	**0.658**	**0.633**	**0.603**	5	1.282	0.772	**0.713**	0.237	1.225

Table 1 continued

Data set	COX-2[a]						HIV-1 RT[b]					
Conformation	A priori		1CX2		6COX		A priori		1VRT		1RT1	
Alignment	Substr.[c]		Surflex[d]	Substr.	Surflex	Substr.[e]	Substr.		Surflex	Substr.	Surflex	Substr.
Orthogonal lattice[l] *LOO*												
SE_C	4	0.628	**0.694**	0.655	0.731	0.710	3	0.598	0.973	**0.938**	1.027	**0.950**
	5	0.618	0.697	0.648	0.717	**0.681**	4	**0.584**	**0.934**	0.954	1.019	0.966
	6	**0.599**	0.704	**0.642**	**0.709**	0.683	5	0.590	0.940	0.977	**1.001**	0.967
q^2	4	0.552	**0.452**	0.518	0.392	0.428	3	0.800	0.471	**0.509**	0.411	**0.496**
	5	0.568	0.451	0.531	0.418	**0.477**	4	**0.812**	**0.518**	0.496	0.426	0.484
	6	**0.595**	0.441	**0.541**	**0.432**	0.475	5	0.810	0.517	0.477	**0.452**	0.488

[a] $N = 304$ unless otherwise indicated

[b] $N = 103$

[c] Ligands were (re)aligned by their core substructures

[d] Alignment generated by Surflex-Dock

[e] $N = 302$

[f] Number of components (latent variables) contributing to each COX-2 statistic

[g] Number of components (latent variables) contributing to each HIV-1 RT statistic

[h] Standard error of regression

[i] Standard error of prediction after 15% response scrambling

[j] Robust q^2 extrapolated to zero perturbation

[k] Sensitivity of q^2 to response scrambling at 15% scrambling; a value above 1.0 indicates a risk of over-fitting [30]

[l] Face-centered cubic lattice at 2.5 Å grid spacing

that makes use of all N observations. An alternative measure of predictivity and statistical robustness that makes use of all available information can be obtained by applying progressive scrambling, in which model stability is assessed by swapping descriptor values between observations with more or less similar biological activity [13, 30]. Comparing the q^2_{LOO} values in Table 1 to the corrected predictivity extrapolated to zero perturbation (Q^2_0 [30]) indicates that a substantial fraction of the a priori 1CX2 model's apparent superiority to the realigned model is actually due to redundancy in the former that is largely absent in the latter. Note that the small idiosyncratic conformational differences induced by docking are themselves a form of 3D perturbation, which will effectively offset simple structural (2D) redundancy.

The differences in predictivity noted here cannot be ascribed to incidental effects of alignment to the lattice. Models were constructed from fields evaluated across face-centered lattice that are essentially orthogonal to the standard cubic CoMFA lattice, and these yielded exactly the same LOO cross-validation trends for the COX-2 models and very similar statistics (Table 1).

The quantitative results seen for the HEPT data set are somewhat different. Realignment of the docked conformations from 1RT1 yielded a much better and more robust model than was seen for the docked alignment itself, whereas realignment reduced the quality of the 1VRT statistics. This is less evident for the model constructed on the orthogonal lattice (q^2_{LOO} 0.524 → 0.475 vs 0.518 → 0.509), implying that the true differential lies somewhere in between. Progressive scrambling, on the other hand, indicates that even the realigned 1VRT model is more robust than the docked alignment that reflects the protein's point of view (Q^2_0 0.472 vs 0.458). Note, too, that all CV statistics for the alignment obtained by docking into apo1RT1 fall below the nominal threshold for statistical significance of q^2, which is 0.40 [60] (Table 1). The three-fold CV result for the realigned 1VRT conformations (q^2_{3x} 0.316) drops below this threshold as well, but the corresponding Q^2_0 (0.472) indicates that this is probably due to an effective reduction in data set size (see above).

It is possible that the docking vagaries described in connection with Figs. 8 and 9 simply represent a 5% failure rate for Surflex-Dock. The structural consistency of the ligands represented, however, argues against this: pyridyl, 4″-oxy, 2″-substituted and 4,4-disubstituted analogs appear out of all proportion to their occurrence in the data set as a whole. In addition, replacing the best-scoring poses with the more orthodox conformations indicated improved all cross-validation statistics appreciably in almost all cases. Unfortunately, it is difficult to assess significance levels for PLS statistics with any precision. That said, it is reassuring that the one area in which CV statistics did not improve was with regard to three-fold cross-validation, precisely where introducing new information should have the most negative impact.

Discussion

Expecting a CoMFA or CoMSIA model to be a "mirror" of the corresponding binding site is intuitively appealing but is dangerously simplistic. Were it the case, one would expect interactions to be very localized and most CoMFA models to be relatively simple, with three or fewer latent variable components. Four to six component models similar to those reported here are more common in practice, which implies that there is a substantial amount of contingent information embedded in them—"Steric bulk *here* is good provided there is an electronegative substituent *there*", for example. The "mirror" paradigm is also hard to reconcile with the predictive successes of Topomer CoMFA, in which alignment rules are completely ad hoc [4]. Finally, one would expect that models constructed from the "authentic" alignments found in crystal structures would consistently outperform models based on more "artificial" alignments. This turns out not to be true, however [61–66].

Some of the unsettling discrepancy between expectation for crystal structures and reality can be attributed to noise introduced by the small incidental perturbations that can arise during crystallization. After all, 1CX2 and 6COX differ appreciably in detail despite being complexes of the same protein and ligand (SC-558). Indeed, the sulfonamide in the PDB structure 1CX2 is nominally rotated 120° with respect to that on 6COX, exchanging one of its oxygens with the NH2 group. This likely reflects the inherent difficulty in distinguishing oxygen from nitrogen within the group itself. In this case, uncertainties in the placement of sulfonamide and hydroxyl hydrogens key to the interaction between inhibitor and protein further complicates matters, as does the difficulty in distinguishing nitrogen from carbon in the imidazole ring of His90. All of these problems are exacerbated by the exposure of both Ser353 and His90 to solvent in the crystal structures.

It seems more likely that the discrepancy reflects systematic changes in the protein around the binding site, as well as variations in binding mode [67–69]. Both effects can profitably be addressed by shifting to ligand-based alignment instead of insisting on a protein-centric frame of reference when using docking or crystal structures to align ligands.

In particular, published CoMFA studies on cyclooxygenase include a study of 24 oxazolone analogs that was complemented by subsequent docking into the 6COX crystal structure [18] using the FlexiDock and AFFINITY docking tools [70]. Inhibitors were aligned by substructure; the docking results were not used for CoMFA. Chakraborti and Thilagavathi used a similar strategy, creating CoMFA

models for 35 1,3-diarylisoindole derivatives active against COX-2, then using FlexX to map the models obtained back into the active site [71]. They found that rigid field fit outperformed straight substructural alignment and that including the calculated molar refractivity (CMR) or some other measure of ligand size in the regression equation improved goodness of fit and predictivity. By analogy, the CoMFA fields obtained by forcing the HEPT analogs into the too-small binding site of 1VRT could be serving as a complicated way to measure molecular volume. If that is indeed what is happening, then unpacking the ligands by realigning them should compromise the model's predictivity—which is exactly what happens for 1VRT (Table 1). It is also consistent with the very low affinities suggested by Surflex-Dock that were noted above.

Datar and Coutinho [15] applied CoMFA to 67 ligands drawn from eight structural families of COX-2 inhibitors, most of which were included here. They found that docking into apo1CX2 using the DOCK and AFFINITY programs yielded somewhat better model statistics that did substructural alignment of relaxed structures to the conformation of SC-558 found in the crystal structure. A similar study on 34 1,5-diarylpyrazoles carried out by Thilagavathi and Chakraborti found that substructural alignment worked much better than docking the inhibitors into 1CX2 using FlexX [16].

HIV-1 RT has also been the subject of previous CoMFA studies. Medina-Franco et al. used AutoDock and apo1VRT to align a series of NNRTIs [39]. Zhou and Madura [40] subsequently applied CoMFA and CoMSIA to 50 TIBO derivatives, another class of NNRTI, in the only analysis directly comparable to this study. They docked their ligands into apo1VRT using Autodock, then compared the CoMFA models for the docked Cartesian alignment with those obtained after realigning the docked ligands by substructure. The docked poses yielded better statistics in their hands, but the authors noted that analogs bearing *any* substituent other than hydrogen at the 8-position on the common core exhibited an altered binding mode. Their result underscores the importance of using central substructures for realignment, thereby minimizing the opportunities for large effects due to fundamental changes in how the ligand core interacts with the protein. Realigning COX-2 inhibitors that dock in "flipped" orientations by their sulfonyl groups, for example, would correspond to such gross reorientation of the protein as to make the field contours obtained uninterpretable.

Conclusion

CoMFA studies carried out on docked ligand poses may serve as useful complements to studies carried out using conformations and Cartesian alignments based solely on substructure, provided a suitable protein target can be identified. They cannot be expected to outperform such a priori alignments in any quantitative sense, but they may provide useful insights nonetheless. This is particularly true where symmetry in the ligands or in the binding site leads to ambiguity in a priori alignment and for ligands with unusual substitution patterns. When docking is used to predict bound conformations, however, the CoMFA model derived from those conformations can be improved statistically and in interpretability by realigning the ligands based on their common substructure—i.e., by putting them into a ligand-based rather than a protein-based frame of reference.

Acknowledgments Drs. Richard Cramer and Brian Masek of Tripos International provided helpful discussions in the work described here. Drs. Ajay Jain (UCSF) and Essam Metwally (Tripos International) have been invaluable collaborators with respect to the Surflex-Dock.

References

1. Cramer RD III, Patterson DE, Bunce JD (1988) J Am Chem Soc 110:5959–5967
2. Klebe G, Abraham U, Mietzner T (1994) J Med Chem 37: 4130–4146
3. Cruciani G, Crivori P, Carrupt P-A, Testa B (2000) J Mol Struct 503:17–30
4. Cramer RD (2003) J Med Chem 46:374–388
5. Beusen DD, Marshall GR (1999) In: Güner O (ed) Pharmacophore perception, development, and use in drug design. International University Line, pp 23–46
6. Perola E, Charifson PS (2004) J Med Chem 47:2499–2510
7. Boström J, Norrby P-O, Liljefors (1998) J Comput Aided Mol Des 12:383–396
8. Boström J (2001) J Comput Aided Mol Design 15:1137–1152
9. Rault S, Bureau R, Pilo JC, Robba M (1992) J Comput Aided Mol Des 6:553–568
10. Calder JA, Wyatt JA, Frenkel DA, Casida JE (1993) J Comput Aided Mol Des 7:45–60
11. Langer T, Hoffmann RD (1998) J Chem Inf Comput Sci 38: 325–330
12. Myers AM, Charifson PS, Owens CE, Kula NS, McPhail AT, Baldessarini RJ, Booth RG, Wyrick SD (1994) J Med Chem 37:4109–4117
13. Clark RD, Leonard JM, Strizhev A (1999) In: Güner O (ed) Pharmacophore perception, development, and use in drug design, International University Line, p 151–169
14. Muegge I, Podlogar BL (2001) Quant Struct Act Relat 20: 215–222
15. Datar PA, Coutinho EC (2004) J Mol Graph Graph Model 23:239–251
16. Thilagavathi R, Chakraborti AK (2005) Internet Elect J Mol Des 4:603–612
17. Gohlke H, Klebe G (2002) J Med Chem 45:4153–4170
18. Kurumbail RG, Steven AM, Gierse JK, McDonald JJ, Stegeman RA, Pak JY, Gildehaus D, Miyashiro JM, Penning TD, Seibert K, Isakson PC, Stallings WC (1996) Nature 384:644–648
19. Marot C, Chavatte P, Lesieur D (2000) Quant Struct-Act Relat 19:127–134

20. Desiraju GR, Sarma JARP, Raveendra D, Gopolakrishnan B, Thilagavathi R, Sobhia ME, Subramanya HS (2001) J Phys Org Chem 14:481–487
21. Chakraborti AK, Thilagavathi R (2003) Bioorg Med Chem 11:3989–3996
22. Kauffmann GW, Jurs PC (2001) J Chem Inf Comput Sci 41:1553–1560
23. Gaudio AC, Montanari A (2002) J Comput Aided Mol Des 16:287–295
24. Rao SN, Stockfish TP (2003) J Chem Inf Comput Sci 43: 1614–1622
25. Liu HX, Zhang RS, Yao XJ, Liu MC, Hu ZD, Fan BT (2004) J Comput Aided Mol Des 18:389–399
26. Baurin N, Mozziconacci J-C, Arnooult E, Chavatte P, Marot C, Morin-Allory L (2004) J Chem Inf Comput Sci 44:276–285
27. Prasanna S, Manivannan E, Chaturvedi SC (2004) QSAR Comb Sci 23:621–628
28. Chavatte P, Yous S, Marot C, Baurin N, Lesieur D (2001) J Med Chem 44:3223–3230
29. Clark RD (2003) J Comput Aided Mol Des 17:265–275
30. Clark RD, Fox PC (2004) J Comput Aided Mol Des 18:563–576
31. De Clerq E (2004) Chem Biodivers 1:44–64
32. De Corte BL (2005) J Med Chem 48:1689–1696
33. Kohlstaedt LA, Wang J, Friedman JM, Rice PA, Steitz TA (1992) Science 256:1783–1790
34. Tanaka H, Takashima H, Ubasawa M, Sekiya K, Nitta I, Baba M, Shigeta S, Walker RT, De Clerq E, Miyasaka T (1992) J Med Chem 35:337–345
35. Tanaka H, Baba M, Hayakawa H, Sakamaki T, Miyasaka T, Ubasawa M, Takashima H, Sekiya K, Nitta I, Shigeta S, Walker RT, Balzarini J, De Clerq E, (1991) J Med Chem 34:349–357
36. Ren J, Esnouf R, Garman E, Somers D, Ross C, Kirby I, Keeling J, Darby G, Jones Y, Stuart D, Stammers D (1995) Nat Struct Biol 2:292–302
37. Hopkins AL, Ren J, Esnouf RM, Willcox BE, Jones EY, Ross C, Miyasaka T, Walker RT, Tanaka H, Stammers DK, Stuart DI (1996) J Med Chem 39:1589–1600
38. Hannongbua S, Nivesanond K, Lawtrakul L, Pungpo P, Wolschann P (2001) J Chem Inf Comput Sci 41:848–855
39. Medina-Franco JL, Rodríguez-Morales S, Juárez-Gordiano C, Hernández-Campos A, Castillo R (2004) J Comput Aided Mol Des 18:345–360
40. Zhou Z, Madura JD (2004) J Chem Inf Comput Sci 44:2167–2178
41. Jalali-Heravi M, Parastar F (2000) J Chem Inf Comput Sci 40:147–154
42. Alves CN, Pinheiro JC, Camargo AJ, Ferreira MMC, da Silva ABF (2000) J Mol Struct 530:39–47
43. Kireev DB, Chrétien JR, Grierson DS, Monneret C (1997) J Med Chem 40:4257–4264
44. Hopkins AL, Ren J, Tanaka H, Baba M, Okamato M, Sturat DI, Stammers DK (1999) J Med Chem 42:4500–4505
45. Luco JM, Ferretti FH (1997) J Chem Inf Comput Sci 37:392–401
46. CONCORD® was created by RS Pearlman, A Rusinko JM Skell, R Balducci at the University of Texas, Austin TX
47. CONCORD® is distributed exclusively by Tripos International, 1699 S. Hanley Rd., St. Louis MO 63144 USA (http://www.tripos.com)
48. Ash S, Cline MA, Homer RW, Hurst T, Smith GB (1997) J Chem Inf Comput Sci 37:71–79
49. SYBYL® is distributed exclusively by Tripos International, 1699 S. Hanley Rd., St. Louis MO 63144 USA (http://www.tripos.com)
50. Gasteiger J, Marsili M (1980) Tetrahedron 36:3219–3288
51. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) Nucleic Acids Res 28:235–242
52. Jain AN (2003) Surflex: fully automaticflexible molecular docking using a molecularsimilarity-based search engine. J Med Chem 46:499–511
53. Surflex-Dock™ is distributed by and by Tripos International, 1699 S. Hanley Rd., St. Louis MO 63144 USA (http://www.tripos.com); and by BioPharmics LLC, 36 Avila Road, San Mateo CA 94402 USA (http://www.biopharmics.com)
54. Jain AN (2007) J Comput Aided Mol Des 21:281–306
55. Cho SJ, Tropsha A (1995) J Med Chem 38:1060–1066
56. Llorens O, Perez JJ, Palomer A, Mauleon D (2002) J Mol Graph Model 20:359–371
57. Soliva R, Almansa C, Kalko SG, Luque FJ, Orozco M (2003) J Med Chem 46:1372–1382
58. Clark RD, Strizhev A, Leonard JM, Blake JF, Matthew JB (2002) J Mol Graph Model 20:281–295
59. CScore™ is distributed exclusively by Tripos International, 1699 S. Hanley Rd., St. Louis MO 63144 USA (http://www.tripos.com)
60. Clark M, Cramer RD III (1993) Quant Struct Activity Rel 12:137–145
61. Waller CL, Oprea TI, Giolitti A, Marshall GR (1993) J Med Chem 36:4152–4160
62. Klebe G, Abraham U (1993) J Med Chem 36:70–80
63. Klebe G, Mietzner T, Weber F (1994) J Comput Aided Mol Des 8:751–778
64. Oprea TI, Waller CL, Marshall GR (1994) J Med Chem 37: 2206–2215
65. DePriest SA, Mayer D, Naylor CB, Marshall GR (1993) J Am Chem Soc 115:5372–5384
66. Martin YC (1998) In: Kubinyi H, Folkers G, Martin YC (eds) 3D QSAR in drug design, vol 3. Kluwer Academic, Dordrecht, pp 3–23
67. Kim KH (2007) J Comput Aided Mol Des 21:63–86
68. De Jonge MR, Koymans LM, Vinkers HM, Daeyaert FF, Heeres J, Lewi PJ, Janssen PA (2005) J Med Chem 48:2176–2183
69. Pouplana R, Lozano JJ, Pérez C, Ruiz J (2002) J Comput Aided Mol Des 16:683–709
70. Desiraju GR, Sarma JARP, Raveendra D, Gopolakrishnan, Thilagavathi R, Sobhia ME, Subramanya HS (2001) J Phys Org Chem 14:481–487
71. Chakraborti AK, Thilagavathi R (2003) Bioorg Med Chem 11:3989–3996