J Comput Aided Mol Des (2006) 20:763–771
DOI 10.1007/s10822-006-9070-2

ORIGINAL PAPER

A marriage made in torsional space: using GALAHAD models to drive pharmacophore multiplet searches

Jennifer K. Shepphird · Robert D. Clark

Received: 20 June 2006 / Accepted: 17 August 2006 / Published online: 3 October 2006

Abstract Pharmacophore multiplets are useful tools for 3D database searching, with the queries used ordinarily being derived from ensembles of random conformations of active ligands. It seems reasonable to expect that their usefulness can be augmented by instead using queries derived from single ligand conformations obtained from aligned ligands. Comparisons of pharmacophore multiplet searching using random conformations with multiplet searching using single conformations derived from GALAHAD (a genetic algorithm with linear assignment for hypermolecular alignment of datasets) models do indeed show that, while query hypotheses based on random conformations are quite effective, hypotheses based on aligned conformations do a better job of discriminating between active and inactive compounds. In particular, the hypothesis created from a neuraminidase inhibitor model was more similar to half of 18 known actives than all but 0.2% of the compounds in a structurally diverse subset of the World Drug Index. Similarly, a model developed from five angiotensin II antagonists yielded hypotheses that placed 65 known antagonists within the top 0.1–1% of decoy databases. The differences in discriminating power ranged from 2 to 20-fold, depending on the protein target and the type of pharmacophore multiplet used.

Keywords GALAHAD · Molecular alignment · Pharmacophore · Pharmacophore multiplets · Tuplets · Virtual screening

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

Abbreviations

ATII	Angiotensin II
CHDPA	Chapman and Hall Dictionary of Pharmaceutical Agents
GA	Genetic algorithm
GALAHAD	Genetic automated molecular overlays
ROC	Receiver Operating Characteristic
WDI	World Drug Index

Introduction

Pharmacophore models consist of an array of specific features important to ligand binding [1, 2], and are generally deduced from information contained in some set of active compounds. A good pharmacophore model, then, captures all or most of the key features common to a given set of actives, along with their spatial distribution in the respective bound conformations [3]. The constituent pharmacophoric features are themselves generalizations of the different ways in which ligands can interact with proteins—typically through ionic, hydrogen bonding or hydrophobic interactions. Because the features are generalized, pharmacophore searches have the ability to ''lead-hop'', i.e., to find compounds that bind receptors in the same way as the known actives but do not share significant substructural similarity with them.

Although this is, in general, a powerful method for finding novel ligands, measuring the effectiveness of a given searching technique can be tricky. An overly general pharmacophore model returns most or all of the actives buried in a mountain of uninformative

inactives. An overly specific pharmacophore model finds some or all actives and few or no inactives, but will miss active molecules that diverge much from the training set, returning only compounds that are structurally similar to the training set. Careful false positive analysis may be necessary in order to gain insight into a receptor-ligand system and to achieve lead hopping.

Classic 3D search queries are based directly on the target pharmacophore. Unfortunately, most binding sites present more potential sites of interaction than can be exploited by any single ligand with good pharmacokinetic properties, so particular pharmacophore features may be found in some ligands but not in all. In addition, some receptor sites contain flexible side chains whose movement changes the shape of the site and the pattern of complementary features presented to ligands somewhat. Such receptor flexibility presents a challenge for structure-based drug design efforts as well as for ligand-based virtual screening.

Fast pharmacophore multiplets (Tuplets [4]) provide an alternative approach to searching that capitalizes on this inherent fuzziness. They capture a molecule's entire pharmacophoric pattern across a broad sampling of conformations or within a single conformation, when the biologically active conformation is known. The full pharmacophoric complement of each ligand is decomposed into its constituent subsets of three- or four-feature multiplets, the distribution of which is encoded in compressed fingerprints, or bitmaps [5]. In their most basic form, multiplets encode two types of information: the pairwise distances between the features (edge lengths; three for a triplet and six for a quartet) and the type of feature found at each vertex (three for a triplet and four for a quartet).

If two ligands share a common pharmacophore, then the multiplets making up that pharmacophore will necessarily appear in the bitmaps generated from the corresponding bound conformations. Comparing bitmaps across ensembles of conformations can reveal the extent to which two molecules can adopt similar pharmacophoric patterns without having to know a priori which multiplets must match: the more matching patterns, the higher the similarity between tuplets. This fuzziness nicely accommodates the inherent fuzziness in protein–ligand interactions alluded to above.

Bound conformations can be extracted from crystal structures of protein : ligand complexes, when enough of these are available. In the absence of such structural information, multiplet hypotheses are commonly based on a set of random conformations generated for each structure in the training set. If the ligands are structurally diverse, there is a good chance that intersection of their bitmaps will be a good approximation to the target pharmacophore. Nonetheless, the fact that a finite number of conformations are used may add an undesirable degree of fuzziness to the hypotheses produced, thereby limiting its ability to discriminate between potential ligands and inactive compounds. Here, we examine the usefulness of the alternative approach of working with conformations derived using a genetic algorithm with linear assignment for hypermolecular alignment of datasets (GALAHAD) [6, 7], a pharmacophore elucidation program that allows for full ligand flexibility while taking strain energy and steric overlap into account. The method used is unique in that it avoids the need for a template structure, which allows for direct and efficient generation of partial-coverage models that include multiple partial match constraints [8, 9].

Datasets

Two different drug classes were analyzed: a set of 18 known neuraminidase inhibitors assembled by Steindl and Langer [10] (Fig. 1), from which six (**1**, **3**, **4**, **6**, **13,** and **14**) were selected as a training set. The training set of five angiotensin II receptor (ATII) antagonists (compounds **19**–**23** in Fig. 2) was taken from the literature [11], whereas the 65 antagonists in the test set were drawn from the MDDR [12]; Fig. 2 includes 13 representative compounds from that test set.

Three different sets of decoys were used:

- A 10,000-compound subset of the World Drug Index (WDI) [13]. A diverse representative subset was selected using OptiSim [14, 15], with molecular similarity and diversity based on the tanimoto similarity of UNITY [4] substructural fingerprints.
- A 1,000-compound diverse representative dataset of proprietary drug candidates (P1K).
- A 400-compound subset drawn from the Chapman and Hall Dictionary of Pharmacological Agents (CHDPA) using OptiSim selection.

Decoys were presumed to be ''false positives'' when plotting receiver operating characteristic (ROC) curves. No attempt was made to identify ''false positives''—i.e., nominally inactive decoys that are, in fact, biologically active. Nor was any attempt made to correct for truly inactive compounds that are so structurally similar to actives that one would want them to be included in any virtual screening program. Hence the recovery statistics presented here are inherently conservative.

All analyses started from 3D structures generated using CONCORD [16].

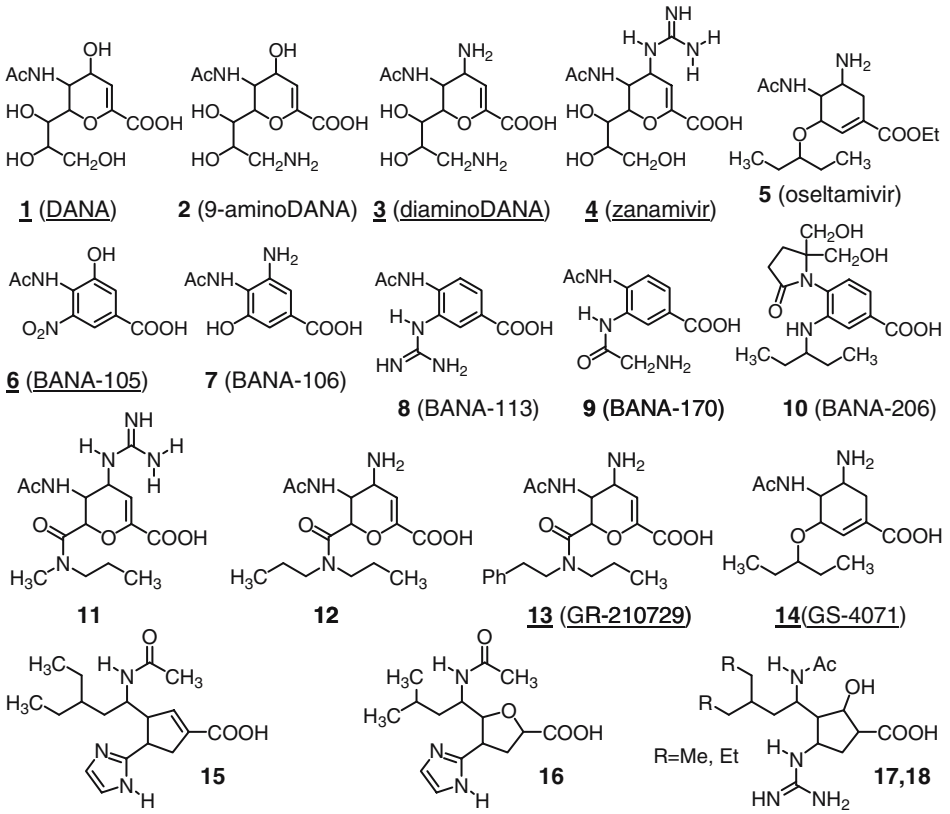

Fig. 1 Neuraminidase inhibitors. Compounds **1**–**10**fall into group I, whereas compounds **11**–**18**fall into group II. Compounds **1**, **3**, **4**, **6**, **13,** and **14** were included in the training set as well as in the test set

Methods

Pharmacophore multiplet bitmaps were generated and analyzed using the Tuplets module in SYBYL 7.2 [4]. Unless otherwise indicated, default configurations and parameters were used, with five feature types—positive nitrogens and negative centers; hydrogen bond donor and acceptor atoms; and hydrophobic centers—taken into consideration and multiplet edge lengths binned at 0.5 Å intervals [5].

Conformations for single-conformer hypotheses were obtained using the GALAHAD program [7–9], which works in two stages. The program first uses a genetic algorithm (GA) to identify a set of ligand conformations that minimizes energy while maximizing pharmacophore multiplet similarity between ligands. Pharmacosteric similarity is maximized at the same time. This operation is carried out in the internal coordinate space, with each ligand flexing simultaneously. The best models are carried forward into the next stage, a rigid-body alignment that overlays the ligands in Cartesian space [9]. Because a multi-objective fitness function is used in the GA [8], multiple models are produced, each of which represents a different trade-off among the competing criteria [17]. Rarely do models achieve optimum values in all parameters simultaneously. For example, the best pharmacophoric match might involve highly strained conformations. Other models might align pharmacophoric features well at the expense of good steric overlap, while yet others may have good steric overlap but fewer common pharmacophoric features. Returning a set of diverse models allows the user to choose the model that corresponds to the most appropriate underlying parameter weights.

Molecular overlays generated by GALAHAD provided the starting point for deriving single-conformer Tuplet hypotheses. Except as otherwise noted, default parameters were used for the GALAHAD runs. Tuplet hypotheses can be generated from two sources of information: pharmacophore queries and sets of active molecules. In this case, pharmacophore queries generated by GALAHAD were not used; rather, each ligand in its aligned conformation contributed to the Tuplet hypothesis.

The GALAHAD-Tuplet workflow employed here is as follows:

1. Assemble a training set of active molecules.
2. Generate models using GALAHAD and select a model exhibiting an acceptable balance between energy, pharmacophoric coherence and pharmacosteric overlap.

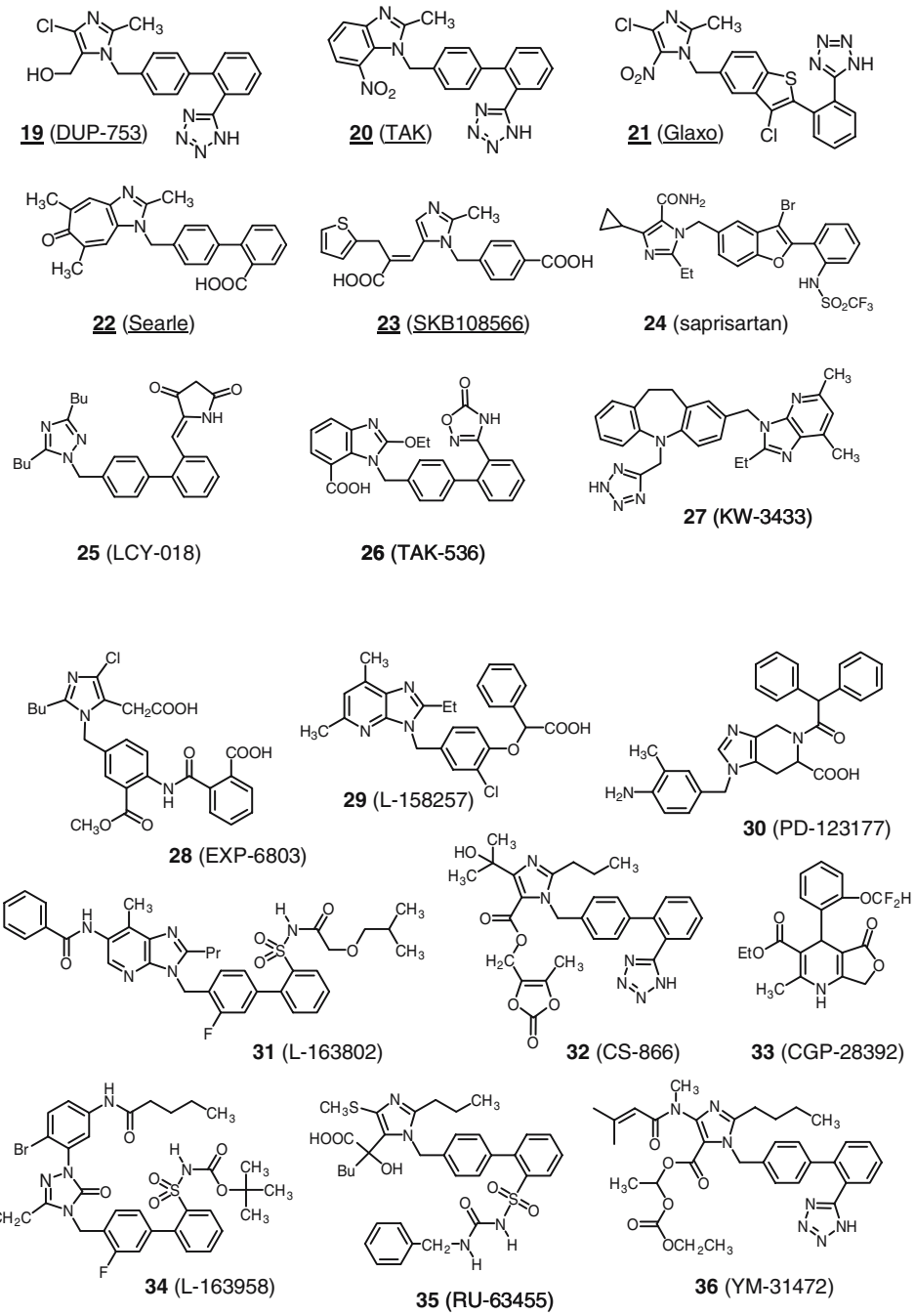

Fig. 2 Representative compounds from the ATII antagonist dataset. The training set was comprised of compounds **19**–**23**

3. Delete the search query features from the model, leaving the aligned ligand structures.
4. Use the ligand conformations from the model to generate single-conformer bitmaps and a single-conformer Tuplet hypothesis.
5. Use bitmaps based on 100 random conformations of each active molecule to generate an alternative, 100-conformer Tuplet hypothesis.
6. Search databases with the Tuplet hypotheses from (4) and (5).
7. Compare the effectiveness of the single- versus 100-conformer hypotheses using ROC plots.

GALAHAD alignments used for Tuplet analysis were selected manually as being representative of the best solutions obtained in each case.

Pharmacophore triplet and quartet bitmaps were used for Tuplet hypothesis generation and database searching. The search databases were multi-conformer Tuplets of the same type as the hypothesis. For

example, if the hypothesis encoded quartets of pharmacophoric features, the search database contained the same type of quartet. Except where otherwise indicated, quartet and triplet hypotheses included the following features: hydrogen bond donor and acceptor atoms, hydrophobic centers, positive nitrogens, and negative centers. Detailed feature definitions are provided in an editable ASCII file, and these include provision for ionization of strong acids and bases, as well as ambiguities due to tautomerization and pK_a's that fall near physiological pH [5].

Each experiment compared two Tuplet hypotheses: one based on a single conformation for each training set compound (from a GALAHAD overlay), and the other based on 100 conformations for each training set compound. Starting with the CONCORD conformation, 100 conformers were generated within the Tuplets program, with torsions about rotatable bonds randomized and steric clashes relieved by the directed tweak method [18].

Receiver operating characteristic curves illustrate the search results by graphically representing the fraction of the known actives recovered (the true positive rate) as a function of the fraction of decoys recovered (false positive rate). Keeping in mind the caveats regarding false positives in virtual screening, ''ideal'' searches would retrieve no false negatives until all actives had been recovered, yielding a vertical line at $x = 0$. Good results have ROC curves close to the upper left corner. For the experiments described here, the area under the curve is high enough (>0.9) that linear ROC plots are not very informative, so semilogarithmic plots are used instead. As a result, random search curves are curved rather than linear and false positive rates of 0 cannot be accommodated. A continuity correction is therefore applied for plotting purposes, with the initial *number* of false positives assumed to be 0.5 compounds rather than zero.

Results

Neuraminidase inhibitors

Both types A and B influenza viruses express neuraminidase on their surface, making it an attractive therapeutic target. The latest commercial class of anti-influenza drugs targeting this enzyme includes zanamivir (**4**) and oseltamivir (**5**). Two modes of interaction are known, differing only in the nature of the ligand moiety that interacts with protein residues Glu 276 and Arg 224. In the first group, the side chain is a glycerol or glycerol-like moiety, as in zanamivir. In the second group, the side chain is bulky and hydrophobic, as in GR217029 (**13**). Glu 276 swings away from the binding pocket to accommodate the bulky side chains of compounds in the latter class (Fig. 3).

The six NA inhibitors selected for the training set included four from class I and two from class II (Fig. 1). GALAHAD yielded the model shown in Fig. 4, from which a single conformer Tuplet hypothesis was generated. The hydrophobic center found in group II inhibitors does not appear in the model query, because it only shows up in two of the six ligands in the training set [9]. It can contribute to the Tuplet hypothesis nonetheless, because many of the multiplets containing it will be large and, consequently, discriminating [5].

The single conformer quartet hypothesis outperformed the multi-conformer hypothesis, especially early in the ROC curve. All 18 known actives were recovered in the top 468, or 4.7% of the database (Fig. 5a). The top 31 compounds (0.3% of the database) contained 50% of the known actives, including five from the training set. Compound **13** (GR217029) was the last recovered from the training set; it ranked 139 (1.4%).

Triplet hypotheses were less discriminating than quartets in both cases (Fig. 5b). Moreover, the difference in performance between the single conformer,

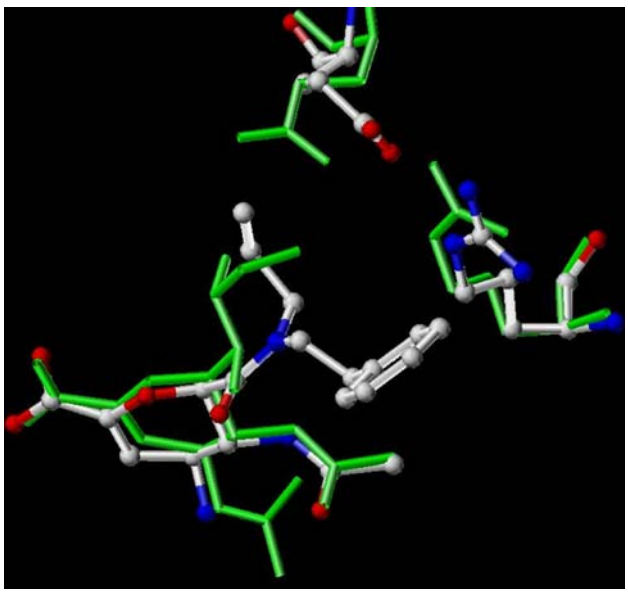

Fig. 3 Several NA inhibitors have bulky hydrophobic substituents that cause Glu 276 (at the top of the picture) to swing away from the binding site and hydrogen bond to Arg 224, which is shown to the *right*. One such compound is **13** (GR-210729; PDB 1bji [19]), shown in *ball-and-stick* and *colored* by atom type. The binding pocket of **4** (zanamivir; PDB 1a4g) is shown in *green* for comparison [19]

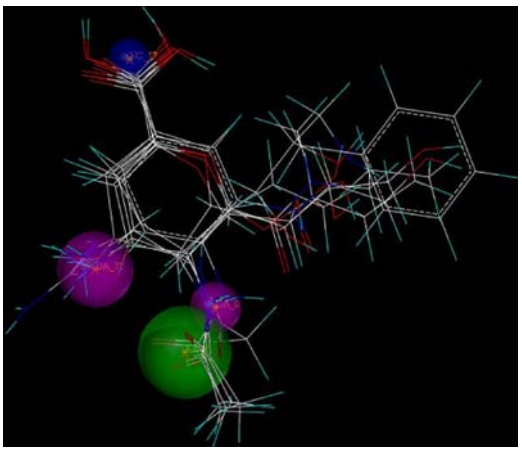

Fig. 4 The GALAHAD model for neuraminidase contains two donor atoms (*purple*), one negative center (*blue*) and one acceptor atom (*green*). The *sphere* sizes indicate query tolerances

GALAHAD-based hypothesis and the one derived from multiple, randomized conformers was smaller—approximately twofold across most of the range in the latter case, as opposed to three to fivefold in the former. This is consistent with the expected higher information content of the more complex multiplet type.

GALAHAD uses a multi-objective fitness function and so returns multiple solutions, with each model representing a somewhat different trade-off between the competing criteria considered—here, pharmacophoric overlap, steric overlap, fit to the query, and total energy. Figure 5c compares the results from four alternative models take from the same GALAHAD run, each of which was superior to the others in some respect. Results for the model shown in Fig. 4 are represented by the solid red line in Fig. 5c. That model was picked out based on its combination of particularly low strain energy and good pharmacophoric overlap. The single-conformer hypothesis based on a more strained model with a somewhat higher pharmacophoric concordance (dashed red line in Fig. 5c) performed equally well. Somewhat counter-intuitively, models with better steric overlap (represented by the orange lines in Fig. 5c) performed even better, particularly for the actives most similar to the training set.

Speed

On an IBM T40p laptop with a 2.0 GHz Intel Pentium processor, Tuplet searches on the 10,000 compounds in the WDI subset took about 9.5 min. For comparison, 3D flexible pharmacophore searches of the same database take between 20 min to many hours depending on the complexity of the query.

Angiotensin II antagonists

Angiotensin II antagonists are used to treat hypertension and congestive heart failure. Many active compounds have been developed since the first non-peptide antagonists were identified in the early 1970s. The five training set compounds were aligned using default GALAHAD parameters, except that the GA population was increased from 50 to 70 and variable indexing was turned on, which allows the linkage between

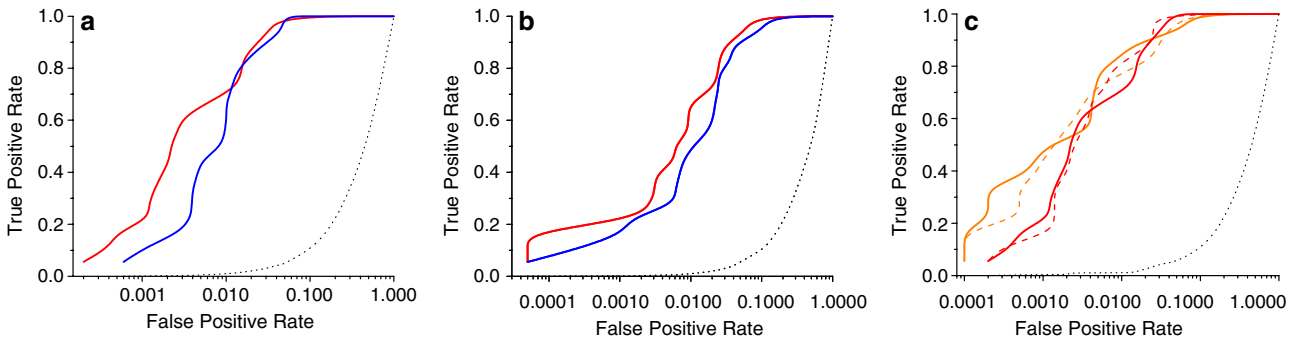

Fig. 5 Receiver operating characteristic curves for neuraminidase Tuplet searches. Six inhibitors provided the training set for the GALAHAD overlays and Tuplet hypotheses. The *solid red curves* are single-conformer hypotheses run against a 10,000-decoy subset of the WDI. The *solid blue curves* correspond to hypotheses based on 100 random conformations. The *dotted curve* represents expected random results. Note that the scale of the *X*-axis is logarithmic; the plot assumes that one false positive is present initially—i.e., that the zeroth compound is one-half of a false positive. (**a**) Hypotheses and searches based on pharmacophore quartets. (**b**) Hypotheses and searches based on pharmacophore triplets. (**c**) Hypotheses and searches based on pharmacophore quartets. The *solid red line* corresponds to a very low-energy model that exhibited good pharmacophoric concordance; this is the same as the single-conformer model in (**a**). The *dashed red line* is for an alternative model with a higher pharmacophoric concordance but more internal strain, whereas the *dashed orange line* is for a low-energy model with paricularly high-steric overlap. The *solid orange line* represents the results obtained when the model ranked highest overall was used

torsions to evolve along with the torsions themselves [8]. The 20 models produced differed somewhat in the number and type of features, and in the conformations and overlay of the molecules, but most contained the key features identified in other studies. These include one or two negative centers, 1–3 acceptor atoms, and several hydrophobic centers. Figure 6 shows the GALAHAD model upon which the Tuplets hypotheses used for subsequent analyses were based.

The results of searches using a single-conformer quartet hypothesis are contrasted with those for a 100-conformer quartet hypothesis in Fig. 7a, with ROC curves shown with respect to both P1K and CHDPA decoy databases. For P1K, the single conformer hypothesis recovered 41 ATII antagonists in the top-scoring 43 compounds and recovered all 65 controls among the top 93. As expected, the 100-conformer hypothesis produced lower recovery rates. Of the top 17 compounds, 13 are known ATII antagonists, and all 65 actives rank in the top 88. Results from searches of CHDPA were similar, though the discrimination seen was somewhat lower. The single conformer tuplet hypothesis obtained from the GALAHAD model recovered most of the actives earlier than did the 100-conformer hypothesis, but the entire active set is recovered about equally well for both the GALAHAD and random-conformer hypotheses.

The ATII pharmacophore model contains two negative centers, important features that appear in most of the actives (58 of 65). Many of the compounds in the proprietary database and CHDPI do not contain negative centers, however. Since these features alone could unduly bias the results, we repeated these analyses after editing negative centers out of the Tuplet configuration file. As can be seen in Fig. 7b, the absence of this feature type had little qualitative effect on the results. For the P1K database, 42 of the top 45 compounds were ATII antagonists, and all 65 appeared in the top scoring 94 compounds for the GALAHAD hypothesis. The multi-conformer hypothesis recovered 17 actives in the top 21 and all 65 antagonists in the top 95.

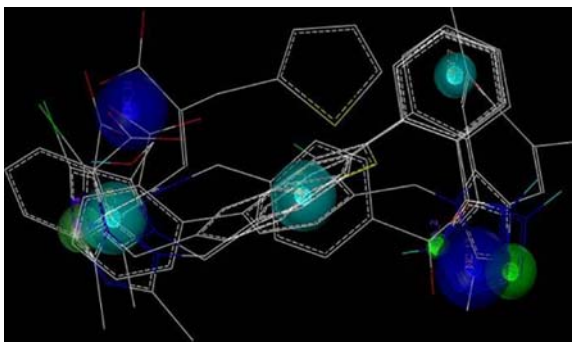

Fig. 6 The GALAHAD model obtained from five compounds in the ATII training set includes three hydrophobes (*light blue*), two negative centers (*dark blue*), and three acceptor atoms (*green*). The *sphere* sizes indicate query tolerances

We repeated the experiments with triplet hypotheses, with the results shown in Fig. 7c. As in quartet searches, ignoring negative centers had relatively little effect (data not shown). Triplet hypotheses produced slightly lower recovery and, again, the single-conformer hypothesis performed more similarly to the 100-conformer hypothesis than for quartets. The triplet differential was about twofold, but in this case the difference in discriminating power between the GALAHAD quartets and those derived from random conformations was much higher—20-fold or more for the best-scoring antagonists. The effect is probably greater here than for the neuraminidase inhibitors because the ligands involved are larger and more flexible.

The difference between the two ways of generating multiplet hypotheses was also explored by comparing the order in which compounds were recovered in each case. The position on the recovery curve (as indicated by the corresponding true positive rate) was very similar for about half of the compounds (Fig. 8a). The fact that the other half exhibit different rank orderings in the single- versus multiple-conformer hypotheses suggests that the two kinds of hypotheses ''see'' similar, but not identical, patterns of pharmacophoric features.

Figure 8b compares the false positive rates at which each antagonist in the test set was recovered. Almost all of the points fall above the 1:1 diagonal, indicating that the GALAHAD hypothesis consistently does a better job of discriminating actives from inactives. The two exceptions are **35** (RU-63455) and **36** (YM-31472), which are highlighted in Fig. 8b. Both have highly congested structures, which suggests that the particular configuration required by the single-conformer hypothesis may be ''out of bounds'' due to steric clashes in these cases.

Discussion

The GALAHAD program uses Tuplet concordances to evaluate the degree of alignment in the internal coordinate space, and produces a UNITY 3D search query. This raises two questions: what is the relationship between the GALAHAD-based Tuplet hypothesis and the ''classic'' one based on the intersection of the ensemble of ligand bitmaps generated from randomized conformations? And, second, how does the

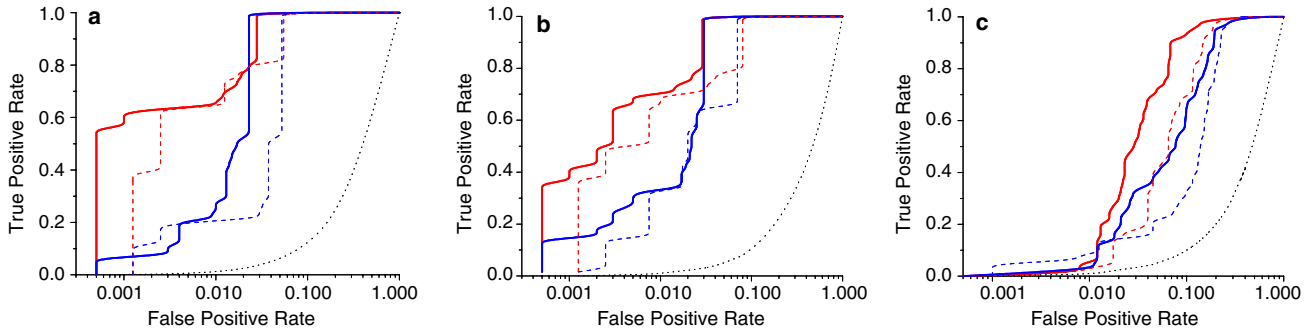

Fig. 7 Receiver operating characteristic curves for AT II Tuplet searches. The results for single conformer hypotheses based on the GALAHAD model are shown in red and those for the 100-conformer hypotheses are shown in *blue*. *Solid lines* indicate search results against the P1K database and *dotted lines* indicate results using CHDPA decoys. The *dotted curve* represents expected random results. The plot assumes that 0.5 false positives are present initially. (**a**) Pharmacophore quartet search results using standard configuration files. (**b**) Pharmacophore quartet search results using configuration files from which negative centers have been removed. (**c**) Pharmacophore *triplet* search results using configuration files from which negative centers have been removed

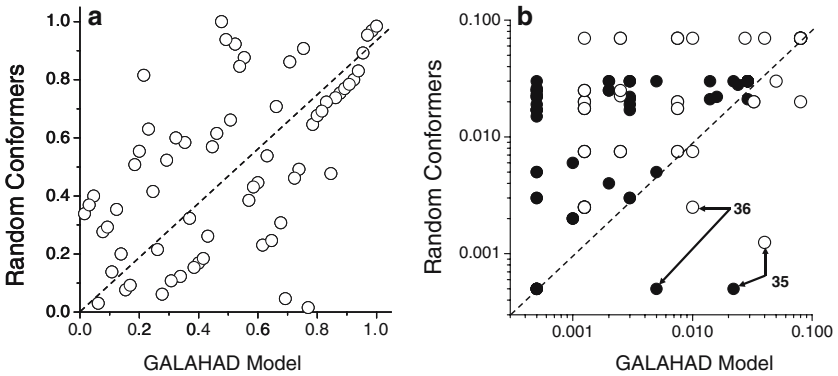

Fig. 8 Recovery statistics for each compound in the AT II test set using the quartet hypothesis based on single conformers from the GALAHAD model versus those using the hypothesis based on 100 random conformers, with negative centers ignored. (**a**) True positive rates representing the rank ordering of the test set. (**b**) False positive rates, indicating how many decoys outscored each active. *Solid circles* represent the results for P1K decoys, whereas *open circles* represent data for the CHDPA subset. The *dashed diagonal* represents the case where the two hypotheses are equally discriminating

GALAHAD Tuplet hypothesis relate to the corresponding UNITY query?

Where exactly one overlay solution exists for a set of compounds and the pharmacosteric and pharmacophoric overlays agree, then the best GALAHAD model will produce the best possible Tuplet hypothesis. When the pharmacosterically and pharmacophorically optimal solutions do not coincide, GALAHAD should perform better since the parameters are more cleanly broken out. Alternative pharmacophore models also present a case where GALAHAD would be superior to the more general Tuplets, where the contributions from multiple possible alignments will tend to average out.

The UNITY query component of a GALAHAD model can, and usually does, include multiple partial match constraints. This, together with the spatial tolerances connected with each feature, confer upon it a degree of fuzziness appropriate for many 3D searching applications. Tuplet queries are, by nature, much fuzzier, however. Hence they have the potential to better accommodate the fuzziness in interaction profiles necessary to compensate for induced fit effects. Single-conformer hypotheses based on GALAHAD models provide a way to exploit that potential while avoiding the noise produced by coincidental concordances between similar multiplets drawn from different ligand conformations.

There are many different ways to search 3D databases, and the best method to use depends on the data on hand as well as the desired results. The ligand-based method we describe here relies on identifying a set of active molecules, aligning them, and then creating and applying fast pharmacophore multiplet search

hypotheses. At the alignment stage, one has flexibility in choosing which GALAHAD overlay best represents binding considerations: pharmacophoric feature match, steric match, minimal ligand strain or some combination thereof. The GALAHAD conformations chosen then provide the starting point for single conformer Tuplet searches. The success of the method will depend, of course, on the robustness of the model chosen.

In both cases examined here, Tuplet searches using hypotheses based on the single conformers in GALAHAD models proved very effective for recovering known actives from among drug-like decoys. This is particularly remarkable given the unusual carboxylic acid bioisosteres represented in the test set (e.g., **25** and **26**), as well as the number of prodrugs it contains (e.g., **32** and **36**). Tuplet hypotheses generated from multiple, random conformers were themselves remarkably effective as well, though somewhat less so than were those based on GALAHAD models.

Acknowledgements Dr. Michael Miller of Pfizer Inc., provided the test set of ATII antagonists used, and Peter Fox of Tripos Inc., provided the subset from the CHDPA.

References

1. Mason JS, Good AC, Martin EJ (2001) Curr Pharm Des 7:567
2. Marshall GR (1993) ESCOM, Leiden, pp 80
3. Martin YC (1993) NIDA Res Monogr 1134:84
4. Tuplets™, SYBYL® and UNITY® are distributed by Tripos Inc., 1699 S. Hanley Rd., St. Louis, MO 63144, USA (http://www.tripos.com)
5. Abrahamian E, Fox PC, Naerum L, Christensen IT, Thogersen H, Clark RDJ (2003) Chem Inf Comput Sci 43:458
6. GALAHAD™ is distributed by Tripos Inc., 1699 S. Hanley Rd., St. Louis, MO 63144, USA (http://www.tripos.com)
7. Clark RD, Abrahamian E, Strizhev A, Wolohan PRN, Abrams C (2005) In: proceedings of the 230th ACS national meeting. COMP, Washington, DC, Abstract No. 137
8. Clark RD, Abrahamian E, Abrams C, Brandt P, Gustavsson A-L, Homan E, N, Metwally E, Richmond NJ, Strizhev A, M, Wirstam M, Wolohan P GALAHAD: a genetic algorithm with linear assignment for hypermolecular alignment of datasets (in preparation)
9. Richmond NJ, Abrams CA, Wolohan PRN, Abraharnian E, Willett P, Clark RD (2006) J Comput Aided Mol Des (in press)
10. Steindl T, Langer TJ (2004) Chem Inf Comput Sci 44:1849–1856
11. Jones G, Willett P, Glen RC (1995) J Comput Aided Mol Des 9:532
12. MDDR is distributed by Prous Scientific (http://www.prous.com)
13. The WDI is distributed by Thomson. The 2005 edition was used here
14. Clark RD (1997) J Chem Inf Comput Sci 37:1181
15. Clark RD, Langton WJ (1998) J Chem Inf Comput Sci 38:1079
16. CONCORD® was developed by Pearlman RS, Rusinko A, Skell JM, Balducci R at the University of Texas, Austin, TX and is available exclusively from Tripos Inc., 1699 S. Hanley Road, St. Louis, MO 63144, USA (http://www.tripos.com)
17. Cottrell SJ, Gillet VJ, Taylor R, Wilton DJ (2004) J Comput Aided Mol Des 18:665
18. Hurst TJ (1994) Chem Inf Comput Sci 34:190
19. Taylor NR, Cleasby A, Singh O, Skarzynski T, Wonacott AJ, Smith PW, Sollis SL, Howes PD, Cherry PC, Bethell R, Colman P, Varghese J (1998) J Med Chem 41:798