J Comput Aided Mol Des (2009) 23:765–771
DOI 10.1007/s10822-008-9227-2

Using a staged multi-objective optimization approach to find selective pharmacophore models

Robert D. Clark · Edmond Abrahamian

Received: 16 January 2008 / Accepted: 30 June 2008 / Published online: 29 July 2008

Abstract It is often difficult to differentiate effectively between related G-protein coupled receptors and their subtypes when doing ligand-based drug design. GALAHAD uses a multi-objective scoring system to generate multiple alignments involving alternative trade-offs between the conflicting desires to minimize internal strain while maximizing pharmacophoric and steric (pharmacomorphic) concordance between ligands. The various overlays obtained can be associated with different subtypes by examination, even when the ligands available do not discriminate completely between receptors and when no specificity information has been used to bias the alignment process. This makes GALAHAD a potentially powerful tool for identifying discriminating models, as is illustrated here using a set of dopaminergic agonists that vary in their D1 vs. D2 receptor selectivity.

Keywords Pharmacophore · Alignment · GASP · GALAHAD · Genetic algorithm · Multi-objective fitness · MOGA

Introduction

G-Protein coupled receptors (GPCRs) are a large and important class of drug development targets, and their agonists and antagonists are often active in pharmacologically useful ways. Unfortunately, they are integral membrane proteins, which makes crystallization of their complexes with ligands difficult and limits the applicability of structure-based drug design tools. In contrast, the rich pharmacology of the area makes GPCR targets attractive candidates for ligand-based development. There the difficulty comes from the fact that most of the GPCRs exist in multiple forms, each with a distinct tissue distribution and a different role to play in the body. The differences between receptor subtypes are often quite small, so a small molecule that binds to one will often bind to other receptor subtypes as well. Such promiscuity can be a good thing for complex disease states that involve simultaneous dysfunction of multiple receptors. More generally, however, it is responsible for undesirable side effects that can restrict the applicability of a drug candidate or block its development altogether.

Sometimes a ligand is available that binds tightly and very specifically to one particular GPCR subtype. If so, that ligand can be used as a template for alignment of other, less specific compounds. More often, however, the ligands available exhibit a range of specificities, or those that are adequately selective do not exhibit a high enough affinity for the primary target to be useful. In either case, alternative ways to align the less discriminating ligands need to be considered, in the hope that different alignments will suit one or another receptor subtype particularly well. The problem is made more complicated if the alignment method used can only make use of interaction features that are shared by all ligands.

GALAHAD [1–3] is a recently developed program designed to generate sets of alternative alignments while making a minimal number of assumptions about how complete pharmacomorphic overlaps need to be. Keys to its operation include the use of fast feature multiplet (Tuplet [1, 4]) technology and employing a multi-objective

R. D. Clark · E. Abrahamian
Tripos International, 1699 South Hanley Road, Saint Louis, MO 63144, USA

R. D. Clark (✉)
Biochemical Infometrics, PO Box 411431, Creve Coeur, MO 63141, USA
e-mail: bclark@bcmetrics.com

fitness function that generates multiple alignments simultaneously, with each alignment representing a different balance between high pharmacomorphic concordances and low energy. This paper describes the application of the program to a set of agonists with varying affinities for the D1 and D2 dopaminergic receptors.

Methods

Like its progenitor GASP [1, 5], GALAHAD seeks to identify a set of ligand conformations that have an optimal combination of pharmacophoric similarity, steric overlap and low strain energy. The two programs both rely on genetic algorithms (GAs) to assure a thorough search of the solution space but they differ in several other respects. Unlike the older program, GALAHAD separates the model building process into two stages—a GA that operates solely in the internal coordinate (torsional) space that is followed by a linear assignment routine which aligns the conformers produced in Cartesian space [3, 6]. Both programs gauge model quality by the same three criteria (strain energy, pharmacophoric overlap, and shape similarity), though these are evaluated somewhat differently in each program. A more fundamental difference lies in how the three criteria are *combined* [7], with GASP using a linear combination of terms as the fitness function and GALAHAD using a true multi-objective (MO) function in which each term is considered independently [8, 9].

Multiplet concordance

Both programs use the Tripos force-field to estimate the torsional and van der Waals energy of the ligands in each model. The way pharmacophoric and steric congruence is measured differs, however. GASP scores each model for how compatible proximal pharmacophoric features in different molecules are and determines pairwise steric overlaps with the template molecule; both terms are evaluated in Cartesian space. GALAHAD, in contrast, seeks to maximize the number of pharmacophoric and steric multiplets shared by the particular ligand conformations that define the model, i.e., to maximize the *concordance* of the corresponding multiplet bitmaps [4, 10, 11].

Concordance is assessed by building a count vector for each class of bitmap,[1] then calculating the sum of squared counts for multiplets present in more than one ligand:

$$\Phi = \sqrt{\frac{1}{n}\sum_{\phi_i > 1} \phi_i^2} \quad (1)$$

$$\Psi = \sqrt{\frac{1}{n}\sum_{\psi_i > 1} \psi_i^2} \quad (2)$$

where n is the number of ligands in the dataset, ϕ_i is the ith element in the pharmacophore count vector, and ψ_i is the ith element in the steric count vector. The sum of squares favors changes that aggregate the counts—e.g., going from 3, 5 and 7 for a particular trio of multiplet counts to 1, 6 and 8 increases the sum of squares from 83 to 100.

Each distinct quartet of features in a particular ligand will always set a bit *somewhere* in the bitmap, but different bits will be set by different conformers because the edge lengths of the corresponding tetrahedra will differ. Therefore singleton multiplets—multiplets that appear in only one ligand's conformer—contribute to neither concordance, since it has to appear *somewhere*.

Every atom in a molecule contributes to the steric overlap considered by GASP, whereas the steric features used in bitmap construction do not fall on every heavy atom [3, 11]. Rather, they are positioned preferentially at terminal groups and bridgehead atoms, then distributed evenly across the rest of the molecule so as to yield about one steric feature for every 2.5 heavy atoms (halogens excepted). The definitions are such that features fall on individual heavy atoms or at the midpoint of a bond between two heavy atoms; those used here are the definitions found in SYBYL 7.4.

Bitmap evaluation takes place within the individual internal coordinate space of each molecule for both types of multiplet and is independent of its position in space with respect to the other molecules making up the model. The absence of a joint frame of reference removes the need for a template molecule and allows the program to efficiently identify queries that include partial-match constraints [12].

Hierarchical multiple objective ranking

A multi-objective function is used for three different purposes in GALAHAD: to assess reproductive fitness, to pick which candidates should survive to the next generation, and to rank models after Cartesian alignment of their constituent ligand conformers. All three functions make use of the *Pareto rank* [9] for each individual model, which is defined as the number of alternative candidates that are better than the model being assessed by *all* criteria; Fig. 1 illustrates how this is evaluated for two competing criteria. The probability of being chosen for reproduction is determined by roulette wheel selection based on:

[1] Just as for bitmaps, the count vector is actually created and manipulated in compressed form.

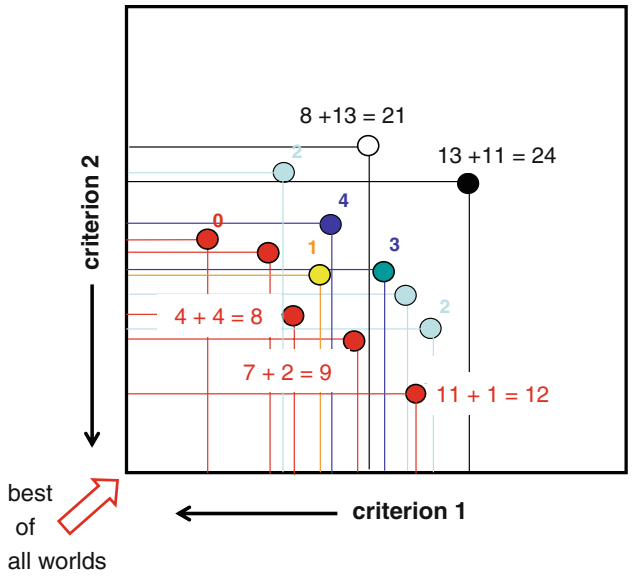

Fig. 1 Schematic illustration of the elements of MOTriage, with each point representing a candidate model; symbols of the same color have the same Pareto rank. The drop lines to each axis embrace other candidates with superior properties by both criteria, which establishes the Pareto ranks shown. Borda tallies are used to break ties; these are calculated as the rank sums shown at selected points

$$p_i \propto \frac{1}{1 + r_i + c_i} \tag{3}$$

where p_i is the probability that model i will be chosen for mating and r_i is the Pareto rank of that model. The c_i term (the number of times the individual has already reproduced in the current generation) is added to reduce the likelihood that any single individual will monopolize reproduction.

Many different individuals in a population can—and generally will—have the same Pareto rank. This is especially so at the *Pareto frontier*, where the rank is 0. Such ties make overall scores difficult to interpret, and can lead to premature convergence when encountered too often in a GA run They can also cause a GA to drift towards models having extreme values of the individual criteria as they progress, which often necessitates specification of exclusion thresholds [9, 13]. To avoid these problems, GALAHAD breaks ties in Pareto rank by applying a hierarchy of *Borda tallies* (sums of ranks across criteria) [14] for final scoring and when determining which models survive intergenerational tournaments.

The Pareto rank of one individual is equal to the number of others in the group being evaluated that are superior to that individual by *all* criteria. This is illustrated for two criteria in Fig. 1, where the (unattainable) best of all worlds can be found at the origin. The number of superior points is simply the number of points falling within the drop lines to the two axes, which is indicated here by color coding: the five models that are superior to all others in at least one respect (i.e., those having a Pareto rank of 0) are colored red, the single model having a Pareto rank of 1 is colored yellow, the three having a Pareto rank of 2 are cyan, and so forth. The individuals at the extremes of the frontier where the Pareto ranks are 0 can be differentiated from those nearer the center of the frontier, where trade-offs are more reasonable, because they tend to have higher (worse) Borda tallies (Fig. 1). In general, the more central models are more desirable; they may not be best in any one respect, but they have pretty good properties over all.

The multi-objective triage (MOTriage) approach used in GALAHAD is actually a nested *hierarchy* of multi-objective functions. Pareto ranking is applied first, then any ties are broken based on the Borda tally across the pharmacophoric and steric multiplet consensus terms. Remaining ties are resolved in favor of the model having the lower overall energy. When the model includes a query, the similarity between the disposition of features within that query and the features in the ligands yields a measure of how well the ligands have been overlaid in Cartesian space. That constitutes a fourth criterion, which is treated on a par with the two multiplet concordances for final model evaluation.

Indeed, nesting the ranks in MOTriage works so well for GALAHAD that piece-wise elitism—guaranteeing that the fittest individual by *each* criterion survives to the next generation—is used by default to minimize the risk of premature convergence.

The dataset

The set of 12 dopaminergic agonists shown in Fig. 2 was taken from Wilcox et al. [15], as were their respective potencies as D1 and D2 agonists. Structures were entered manually in SYBYL 7.4 [1], standardized using CONCORD [16] and minimized using the MMFF94s force field [17]. Both stereoisomers were generated at epimeric nitrogens and, for 6,7-ADTN, at the chiral carbon. The ensemble of ligands was brought into SYBYL as a directory of multi-mol2 files, with each multi-mol2 file containing the alternative configurations of one ligand. GALAHAD was allowed to flex all non-terminal single bonds not in rings and to choose among the epimeric alternatives provided.

The unprotonated ligands shown in Fig. 2 were used to avoid having net charges complicate energy calculations. Note, however, that protonation state has no effect on the recognition of the tertiary nitrogens as positive centers and as hydrogen bond donors, since the macro definitions used anticipate that they will be protonated at physiological pH [4].

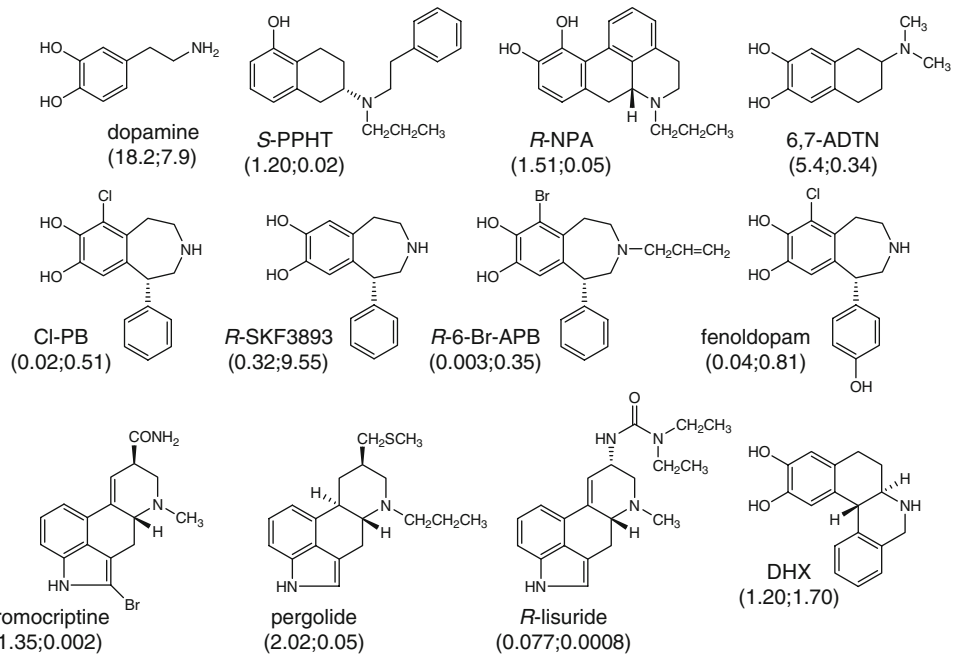

Fig. 2 Structures of the dopaminergic agonists used in the studies described here. The values shown in parentheses are K_L (in μM) with respect to D1 and D2, respectively (Wilcox et al. [15])

Results

GALAHAD was run for 100 generations with a population size of 70 and a tournament pool size of 210. Pharmacophore and steric quartets were used to evaluate overlap during the GA run. At least seven ligands had to contribute to any given consensus feature for it to be included in the model query; lower levels of stringency tend to produce partial match constraints that are too weak to be useful. Default values were used for other settings.

The two models with the best overall score were focused on NPA and ADTN, both of which are selective for the D2 receptor. The third-best model was mixed, but the central ligands in the fourth—Model_004—were fenolopam and SKF38393, which are selective for D1. Model_001 and Model_004 are shown in Figs. 3, 4, along with the partial-match UNITY search queries generated from each overlay.

The pharmacophore queries are very similar in the two subtype models, as were the pharmacophores generated manually in earlier analyses [15]. The queries are quite symmetrical, however, and some ligands "hit" the vicinal donor/acceptor pairs differently in the two models. Recall, too, that the program considers *all* inter-ligand feature correspondences in trying to optimize the overlay, not just those satisfying the minimum level of generality (here, 7 out of 12) required for ultimate inclusion in the query derived from that overlay. These kinds of secondary interactions lead to the difference in shape of the two models, and the difference in shape reflects how otherwise degenerate pharmacophoric features like *ortho*-dihydroxyphenyl (catecholic) groups fit into them.

Pharmacomorphic differences between the models can be further delineated by considering those ligands that did *not* satisfy the model query constraints, the structures of which have been omitted from Fig. 3 but are included in Fig. 4. The features in these ligands participate in alignment and query generation [3], and so are returned as part of the GALAHAD model. All are highlighted in purple at the end of a GALAHAD run, but here each has been assigned a different color to illustrate how it fits into each model.

Three of the "missing" molecules—pergolide, bromocriptine and lisuride—are potent agonists. They fail to satisfy the model queries because they lack the distinctive catecholic substructure feature shared by the other agonists, even when the more D1-specific benzazepines are omitted from the analysis (data not shown). A key difference between the two models with respect to these ligands lies in how their indole nitrogen donor atoms are positioned. In Model_001, they are overlaid with the phenolic groups but do not "hit" the corresponding query because they fail to satisfy the 2 out of 3 partial match constraint, regardless of whether the model conformations are allowed to flex. In Model_004, on the other hand, the indole donor atoms are positioned near the distal *p*-phenolic group of fenoldopam. Were they a little closer together and the consensus threshold for identifying shared features set lower than the value of 7 used here, GALAHAD would have generated an additional partial match feature in this region of the model; requiring only 4 out of 12 features would likely not yield useful models or search queries, however.

This more complete—albeit more complex—view of the system suggests that the model favoring D1-specific agonists (Model_004; b in Figs. 3, 4) takes up a considerably

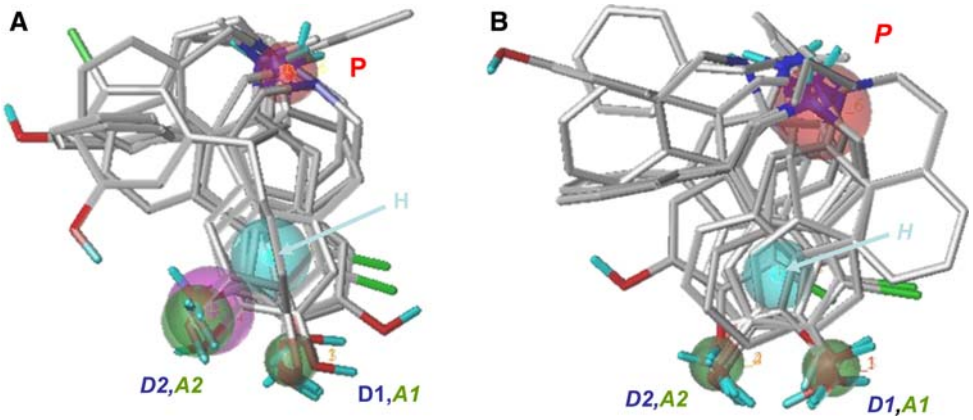

Fig. 3 GALAHAD overlays and UNITY queries for dopaminergic agonists. Cyan indicates hydrophobic centers in the query, red features are positive nitrogens, green indicates hydrogen bond acceptors, and magenta indicates hydrogen bond donors. Molecules failing to "hit" the query are omitted here for clarity. (**a**) Model_001, which favors agonists specific for the D2 receptor. One donor atom and the two acceptor atoms constitute a partial match constraint, with two of the three required to satisfy the constraint. (**b**) Model_004, which favors agonists more specific for the D1 receptor. All six features are included in a single partial match constraint, with a minimum of four features required to match

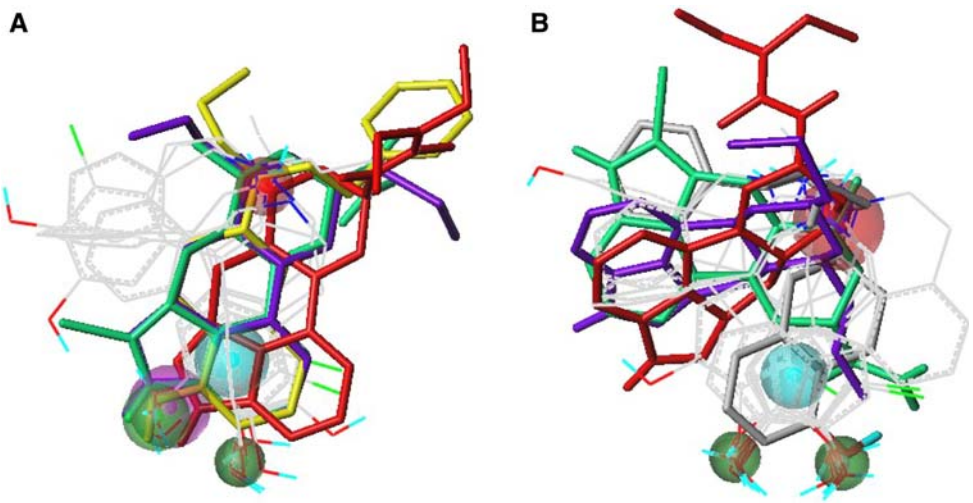

Fig. 4 Pharmacomorphic differences between discriminating GALAHAD models. Ligands that do not "hit" the query in at least one of the models are drawn as capped sticks, whereas other ligands are shown as frames. Pergolide is highlighted in purple, bromocriptine is highlighted in green-blue and lisuride is highlighted in red. (**a**) Model_001, which favors D2-specific agonists. PPHT failed to "hit" this query and is highlighted in yellow. (**b**) Model_004, which favors D1-specific agonists. PPHT, which satisfied the model query, is colored by atom type

larger volume than does the one that favors more D2-specific agonists (Model_001; a in Figs. 3, 4).

Interestingly, a single epimer was selected by GALAHAD for all but two ligands: lisuride came out as the *N*(R) epimer in Model_001 but as *N*(S) in Model_004; the opposite result was seen for *S*-PPHT. In all cases, the preferred configurations matched those specified by Wilcox et al. [15] It is also worth noting that the model that favors D2-selective agonists had a better MOTriage score than did the model that favors D1, a result consistent with the difference in affinity seen for the best agonist in each case: $K_L = 0.8$ nM for D2 (lisuride) and 3 nM for D1 (Cl-PB).

For each ligand in a model, GALAHAD returns a score indicating how much that ligand contributed to each different part of the model's score. These component scores are not in general linear free energy descriptors, but they are expected to be monotonic with respect to free energies. Hence their *ranks* are expected to correlate linearly with the *ranks* of the corresponding affinities [18]. This means that the ranks of the differences in scores between models can reasonably be expected to correlate with observed

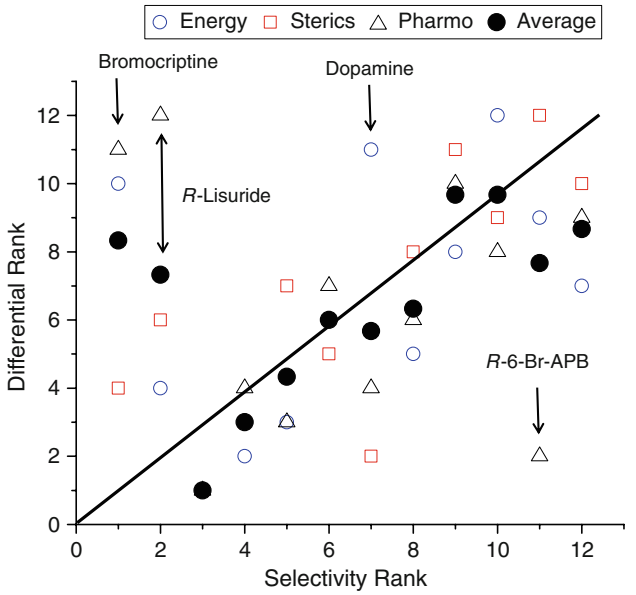

Fig. 5 Rank correlations between selectivity and differences in the goodness-of-fit scores for individual ligands calculated for Model_001 and Model_004. The line shown represents a perfect 1:1 correlation. (○) Inverse rank of the difference in energy. (□) Rank of the difference in steric quartet concordance. (△) Rank of the difference in pharmacophore quartet concordance. (●) Rank average across all three criteria

selectivity ranks. This expectation is largely realized for the dataset considered here; Fig. 5 shows the ranks of the differences in energy and pharmacomorphic concordances between Model_001 and Model_004 as functions of the rank of the specificity calculated from the ratio of the affinities (K_L^{D1}/K_L^{D2}). Note that the ranks of the ratios are the same as the rank of the differences in affinity on a logarithmic scale that is linear in free energy. It also shows the average of the ranks across all three criteria as a function of the specificity rank. The energy rank is inverted, since it is expected to correlate in the opposite sense to that of the other criteria.

Two indoles—bromocriptine and *R*-lisuride—show major positive deviations from the general trend for all criteria, whereas *R*-6-Br-APB shows a substantial negative deviation in terms of steric concordance. Dopamine itself—which lies near the median of selectivity—exhibits relatively large positive *and* negative deviations for different criteria, which cancel out in the average. Dopamine is the native ligand, so it is not too surprising that evolution has seen to it that the balance of interactions between the natural effector and its receptors is exceptionally fine-tuned. The deviations for bromocriptine and *R*-lisuride are harder to rationalize, but they bind so tightly to D2—with 25 and 10 times the affinity seen for the next best ligand—that they may simply fall outside the applicability domain of the SAR. Then, too, their structural similarity belies their observed 17.5-fold difference in affinity for D1.

Discussion

The approach used here to identify subtype-specific alignments and pharmacophores is a form of unsupervised pattern recognition. One could instead build one model from ligands that are relatively more potent at the D1 receptor and a second model from ligands selective for the D2 subtype. Doing so would ignore the fact that all of the ligands bind to both receptors, however, and runs the risk of discarding useful information. In this particular case, such a supervised strategy proves uninformative because the D1-specific agonists are all benzazepines, whose overlay is trivial. Providing multiple ring configurations doesn't help the situation; the output is a simple overlay of the lowest energy ring configuration. Restricting oneself to the D2-specific agonists produces both types of models, since both classes of activity are present. The models obtained are somewhat lower in quality, however, since they are constructed using less information. Training sets comprised only of agonists exhibiting intermediate selectivity, on the other hand, are likely to be underdetermined and yield fuzzy models representing a blend of the subtype pharmacomorphs. The unsupervised approach allows *all* ligands to contribute to *all* models, making it possible to generate an ensemble of models spanning the alternative trade-offs represented by different binding mode overlays.

The only differences in ring structure considered here involved stereochemistry. Including alternative ring conformations but keeping other parameters the same yielded the two models shown in Fig. 6. These models are very similar to the pair produced with single MMFF94-refined CONCORD ring configurations. The displacement of the positive nitrogen for D1 above the plane defined by the other model features nicely mirrors that seen in the manual pharmacophore analysis [15].

GASP is not designed to accommodate the inherent ambiguity represented by the full set of mixed-specificity ligands. It is suitable for the alternative supervised approach outlined above, where it gave identical results to those produced by GALAHAD for the D1-selective benzazepines. GASP had difficulty with the (relatively) D2-specific ligand subset, however; a single class of model was generated in which the catechol rings were precisely superimposed. Unfortunately, such a crisp overlay results in the distal donor atom/acceptor pair corresponding to GALAHAD's positive nitrogen being either missing altogether or having impractically large tolerances—in excess of 1.8 Å.

Inhibitors of serotonin and norepinephrine reuptake have yielded qualitatively similar results, as have proprietary datasets. It should be noted, however, that such analyses are less likely to be fruitful in cases where most or all of the ligands in question are very flexible.

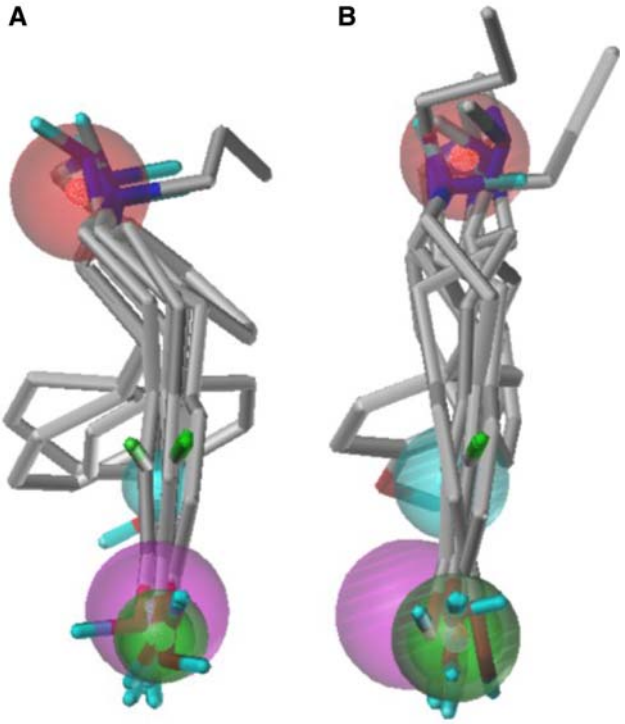

Fig. 6 Edge-on view of GALAHAD overlays and UNITY queries for dopaminergic agonists when multiple alicyclic ring conformations are considered. The feature coloring scheme used is the same as that as that for Fig. 3. (**a**) Model favoring agonists selective for the D1 receptor. (**b**) Model favoring agonists selective for the D2 receptor

Conclusion

Complex partial match queries are often needed to differentiate effectively between related G-protein coupled receptors and their subtypes. The multi-objective fitness and scoring functions used in the GALAHAD program lead to multiple pharmacomorphic models, each of which represents a distinct trade-off between internal strain, pharmacophoric overlap and steric overlap. The different models produced by this unsupervised alignment procedure can then be examined to see if ligands favoring a particular subtype fit one model particularly well. This approach can be productive even when the ligands available are less discriminating than one might wish. The dopaminergic agonists considered here show that GALAHAD can be a useful tool for identifying subtype-specific pharmacomorphic models when used in this way. Indeed, the results suggest that different receptor subtypes may differentiate between ligands because their 3D structure "chooses" one particular tradeoff between energy and steric and pharmacophoric complementarity when several different semi-degenerate alternatives are energetically accessible.

Acknowledgment The suggestions provided by anonymous reviewers contributed significantly to the clarity and readability of this paper, and the authors appreciate their thoughtful criticisms.

References

1. GALAHAD®, Tuplets™, GASP™ and SYBYL® are available from Tripos International, 1699 S. Hanley Rd., St. Louis, MO 63144, USA. http://www.tripos.com
2. Clark RD, Abrahamian E, Strizhev A (2007) US Patent Application 20070143030. Accessible at http://appft1.uspto.gov/netahtml/PTO/srchnum.html
3. Richmond NJ, Abrams CA, Wolohan PRN, Abrahamian A, Willett P, Clark RD (2006) J Comput Aided Mol Des 20:567–587. doi:10.1007/s10822-006-9082-y
4. Abrahamian E, Fox PC, Nærum L, Christensen IT, Thøgersen H, Clark RD (2003) J Chem Inf Comput Sci 43:458–468. doi:10.1021/ci025595r
5. Jones G, Willett P, Glen RC (1995) J Comput Aided Mol Des 9:532–549. doi:10.1007/BF00124324
6. Richmond NJ, Willett P, Clark RD (2004) J Mol Graph Model 23:199–209. doi:10.1016/j.jmgm.2004.04.004
7. Gillet and coworkers have explored different approaches to incorporating multi-objective fitness and partial match capability into GASP. For details, see: Cottrell SJ, Gillet VJ, Taylor R (2006) J Comput-Aided Mol Des 20:735–749. doi:10.1007/s10822-004-5523-7
8. Gillet VJ, Khatib W, Willett P, Fleming PJ, Green DVS (2002) J Chem Inf Comput Sci 42:375–385. doi:10.1021/ci010375j
9. Coello CAC, Van Veldhuizen DA, Lamont GB (2007) Evolutionary algorithms for solving multi-objective problems. Springer, New York
10. Clark RD, Fox PC, Abrahamian E (2005) In: Alvarez J, Shoichet B (eds) Virtual screening in drug discovery. CRC Press, Boca Raton, pp 207–225
11. Abrahamian A, Clark RD, Fox P, Metwally E (2007) US Patent No. 7,212,951
12. Shepphird JK, Clark RD (2006) J Comput Aided Mol Des 20:763–771. doi:10.1007/s10822-006-9070-2
13. Cottrell SJ, Gillet VJ, Taylor R, Wilton DJ (2004) J Comput Aided Mol Des 18:665–682. doi:10.1007/s10822-004-5523-7
14. Cvetković D, Coello CAC (2005) In: Jin Y (ed) Knowledge incorporation in evolutionary computation. Springer, New York, pp 479–503
15. Wilcox RE, Tseng T, Brusniak M-YK, Ginsburg B, Pearlman RS, Teeter M et al (1998) J Med Chem 41:4385–4399. doi:10.1021/jm9800292
16. CONCORD® was developed by R. S. Pearlman, A. Rusinko, J. M. Skell and R. Baducci at the University of Texas, Austin, and is distributed exclusively by Tripos International, 1699 S. Hanley Rd., St. Louis, MO 63144
17. Halgren TA (1999) J Comput Chem 20:730–748. doi:10.1002/(SICI)1096-987X(199905)20:7<730::AID-JCC8>3.0.CO;2-T
18. Duewer DL, Clark RD (1991) J Chemometr 5:503–521. doi:10.1002/cem.1180050604