Journal of Computer-Aided Molecular Design **18:** 523–527, 2004.
DOI 10.1007/s10822-004-4065-3

Descriptors you can count on? Normalized and filtered pharmacophore descriptors for virtual screening

Andrew C. Good[a,*], Sung-Jin Cho[b] & Jonathan S. Mason[c]
[a]*Bristol-Myers Squibb, 5 Research Parkway, Wallingford, CT 06492, USA;* [b]*One Amgen Center Driver, Thousand Oaks, CA 91320, USA;* [c]*Pfizer Central Research, Ramsgate Road, Sandwich, Kent CT13 9NJ, UK*

Received 30 April 2004; accepted in revised form 8 September 2004

Key words: conformational flexibility, fingerprint, normalization, pharmacophores, virtual screening

Summary

The three-dimensional (3D) binary pharmacophore fingerprints find wide application as descriptors in applications ranging from virtual screening through library design. While the 3D content they capture is an intuitively attractive feature of such measures, maximizing their signal to noise ratio has proven to be a tricky balancing act. This issue surfaces primarily due to the potential of such fingerprints to create an explosion of pharmacophores as molecular complexity and flexibility increases. In this article, we describe a modification to the fingerprint generation process that normalizes pharmacophore occurrence frequency by the conformational ensemble size used to derive the descriptor. By including pharmacophore frequency and conformational count, the importance of a given pharmacophore is weighted by the probability of its existence within a given conformational ensemble, rather than treating each pharmacophore equally. In addition, a number of filters have been added to permit the removal of unwanted pharmacophores from the descriptor set. These filters are based on pharmacophore composition (e.g. permutations made up primarily of lipophilic and/or aromatic centers), and size (pharmacophore perimeter length relative to the largest perimeter length found in the molecule). The highly uneven nature of pharmacophore distributions across the conformational ensemble used to generate them is highlighted, as are enrichment comparisons with their binary fingerprint peers. In addition, the limitations in descriptor comparison validation are highlighted as an illustration of the need for more extensive validation experiments.

Introduction

Pharmacophores are generally defined as a critical three-dimensional (3D) geometric arrangement of molecular features or fragments forming a necessary but not sufficient condition for biological activity. While these descriptors find wide use in computer-aided molecular design (CAMD) [1, 2], in this article we are primarily interested in their application as whole molecule measures for molecular similarity calculations [3–7]. Such descriptors typically measure the distances between pairs [3], triplets [6] and/or quartets [7] of key interaction centers, building the resulting pharmacophore ensemble into a fingerprint. Of particular interest in these studies are fingerprints derived via the combination of conformational ensembles [6, 7] (Figure 1). The general structure of such fingerprints has typically been formulated to be binary in nature. Since only presence or absence of a given feature can be determined using such measures, the result is an implicit assumption that an even distribution exists with respect to descriptor bin occupation. The lack of such a distribution in the context of fingerprints derived from combinatorial library pharmacophore ensembles has been highlighted previously [8]. A

* To whom correspondence should be addressed. E-mail: andrew.good@bms.com

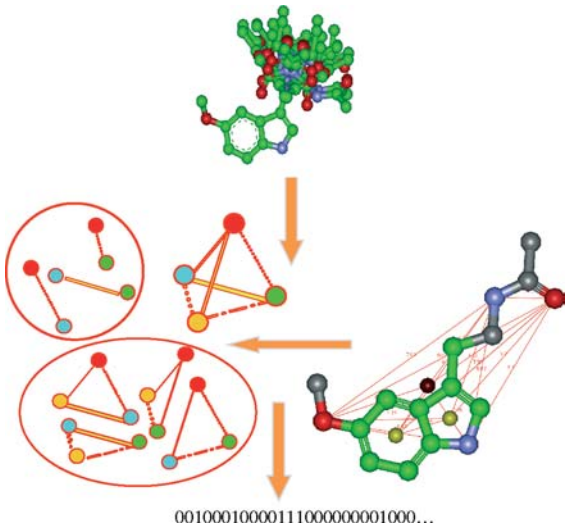

Figure 1. Schematic of binary pharmacophore fingerprint construction.

similar non-uniformity was expected to hold true for pharmacophore distributions across a conformational ensemble, producing a significant degradation of the signal to noise ratio in the resulting fingerprint. This is exacerbated by the fact that extremely rough forms of conformational analysis are typically applied to such descriptor calculations. This produces large increases in the noise content of the resulting fingerprints. With this in mind, efforts were also made to devise methods for their removal. The techniques used and some preliminary results obtained using the resultant descriptors are detailed below.

Methodology

Using Chem-X [9], pharmacophore features deemed likely to be significant in drug–receptor interactions are automatically identified for each molecule. The six key features used in this work are hydrogen bond donors; hydrogen bond acceptors; acidic centers (negatively charged at physiological pH 7); basic centers (positively charged at pH 7); hydrophobic regions; and aromatic ring centroids. Conformational sampling used in fingerprint generation employed Chem-Diverse 'on-the-fly' rule-based conformational generation [1, 9] at search time, with distances between pharmacophoric features for all the evaluated conformers calculated and divided into 6 distance bin ranges [7]. All valid combinations of features and distance ranges define the potential pharmacophores, the presence of which in any conformer of a molecule sets a bit in the pharmacophore fingerprint. These procedures are derived directly from 4 center binary pharmacophore fingerprint generation techniques applied in the earlier work of Mason et al. [7], where additional details regarding calculation procedures are to be found. The pharmacophore generation scripts used in binary fingerprint generation have been modified to permit the inclusion of pharmacophore and conformational count. The resultant data is then used to normalize each pharmacophore contribution by dividing its frequency of occurrence by the conformational ensemble count. For example, a pharmacophore found once in 50 of 100 conformers would score 0.5. The resulting histogram descriptor thus contains a series of floating point values rather than bits associated with each pharmacophore found. This is done as a post processing calculation on the Chem-Diverse phamacophore data files, as is the filtering of 'redundant' pharmacophores. Redundancy is currently based on the removal of low-content pharmacophores [10]. For these calculations LLLL, RRRR, LLL*, RRR*, RRLL, BB**, CC** were removed, where L = lipophilic, R = aromatic, B = base and C = acid. The multi lipophilic/aromatic pharmacophore combinations were removed for their low information content and perceived universality in molecular pharmacophore space. The multi acid / base combinations were removed to mitigate a pharmacophore definition issue in which functionalities such as guanidines and tetrazoles contribute multiple basic / acidic points to the pharmacophore ensemble. These points result in a degree of substructure bias based on multiple acidic / basic points in close proximity to each other. Since pharmacophore-based descriptors are designed primarily with scaffold hopping in mind, such bias is generally best avoided. Removal of these points creates a more level playing field when comparing functionality contributing a varying number of such pharmacophore points. In addition, we generally prefer to avoid leads already carrying multiple charge centers, since they tend to be associated with poor permeability. Similarly a perimeter cutoff of 0.5 was used to remove pharmacophores too small to represent a 'significant' portion of the molecule (the term 'significant' is of

course open to interpretation). Perimeter here is defined as the sum of atom pair distances in a given pharmacophore. The cutoff is classified as the perimeter of the pharmacophore of interest divided by the largest perimeter found in the full pharmacophore ensemble of the current molecule. The resulting descriptors are thus designed to contain pharmacophores with the highest information content, weighted toward those with the highest frequency of occurrence during conformational search.

Results

Two tests were undertaken to probe the behavior of the new pharmacophore descriptors. In the first, the distributions of pharmacophore occurrence frequencies and effects of filtering were analyzed. In the second, enrichment test comparisons were undertaken to compare descriptor performance. In both cases data was abstracted from a collated list of 38 Factor X (FXa) inhibitors, divided into 19 chemotypes based on the novelty (primarily defined around project chemists perspectives of target IP space) of key substructure elements (P1 base mimic and P1–P4 linker core).

(1) To gain insight into the potential effects of normalization and filtering, an analysis of data set pharmacophore frequency distributions was undertaken. For each molecule, normalized pharmacophore histograms were created as described above. The resulting histograms were then collated for all pharmacophores in the data set to create an ensemble distribution of normalized pharmacophore scores (153,863 pharmacophores). The results are shown in Figure 2. In addition, the effect of pharmacophore filtering was studied for the full data set by determining how many pharmacophores were removed through their application. This number was found to equal 71,120, or around ~46% of the pharmacophore total, with ~58% of the pharmacophore removed coming from the perimeter cutoff.

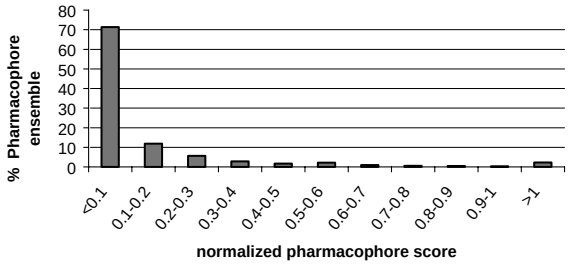

Figure 2. Pharmacophore frequency distribution for the FXa data set.

(2) To test the behavior of the new descriptor in virtual screening calculations, the active compound collection was combined with a noise data set, comprising 9524 molecules used to simulate inactive compounds. These molecules were selected from a 10,000 compound random selection of our in-house database with a conformational flexibility distribution that mirrored the parent data set. All molecules with more than 15 rotatable bonds and Concord [11] build failures were removed. Enrichment calculations were designed to provide an overall measure of a descriptor's ability to pick out novel actives from a hit list, in keeping with the primary goal of virtual screening. This was accomplished by calculating the similarity of each molecule in turn versus all other molecules in the data set together with the noise compounds. The enrichment rate was then determined based on the number of new chemotypes found (members of the template molecule's original chemotype ignored to prevent active analog bias). The results for all molecules in each chemotype were then averaged to provide an indication of the overall ability to enrich given a single chemotype to work with. These chemotype enrichment results were then summed and divided by the total chemotype count to produce a final averaged value of chemotype jumping ability. For the virtual screening similarity calculations the binary fingerprints results are again based on the work of Mason et al. [7] using Equation 1,

$$\frac{O}{w(T-O)+w(P-O)+O} \tag{1}$$

where O = number of pharmacophores in common, T = total number of pharmacophores found in the template molecule, P = total number of pharmacophores found in probe molecule and w = weighting factor, which varies as follows: 2 centers = 1.0, 3 centers = 0.5, 4 centers = 0.11. For the normalized and filtered pharmacophore descriptors simple pharmacophore overlap (numerator term) was used, with no denominator terms added. Many other similarity formulae could have been chosen for these calculations [12]. Equation 1

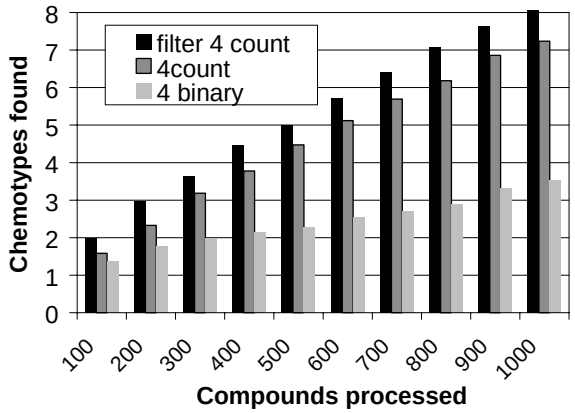

Figure 3. Enrichment results comparing normalized (4 count), normalized / filtered (filter 4 count) and binary (4 binary) 4 center pharmacophore descriptors for the FXa data set.

has been extensively optimized for 4 center pharmacophore in-house, however, and thus was deemed to provide the most internally relevant control. The simple numerator term was chosen to provide a comparison in performance based on descriptor modification that was not obscured by complex coefficient construction. Further optimization might well be possible, however, if a denominator term reflective of molecule size / complexity differences were also constructed. The results of these studies are shown in Figure 3.

Discussion

Results from the first study illustrate that, as expected, pharmacophore frequency distribution is highly uneven (Figure 2). For the FXa data set $>70\%$ of pharmacophores occur in $<10\%$ of the conformers of molecules from which they are derived, highlighting the reasoning behind normalization. Filtering is found to exhibit similar potential as a method for removing noise, with $>40\%$ of pharmacophores removed from the data set total after application.

Study two confirms that for flexible inhibitor templates in the FXa data set, both normalization and filtering are able to increase screening enrichment levels (Figure 3). Calculations involving the binary pharmacophore descriptors were undertaken using Equation 1. This formula's innate complexity is designed to mitigate the high sensitivity of such calculations to small changes in molecular functionality (the number of pharmacophores varies as $n(n-1)(n-2)(n-3)/4!$, the number of pharmacophore centers $\times$ the number of conformers generated in a given molecule). In contrast, the new pharmacophore descriptors employ a simple numerator measure for similarity, relying instead on the descriptor modifications to increase signal. Normalization clearly has the desired effect of improving enrichment for this test case, though it should be noted that no attempt has been made to improve this further through equation experimentation. It is also worth noting that no attempt has been made thus far to optimize the filtering procedure. The perimeter ratio cutoff and low-content pharmacophore set selection criteria were based strictly on structural intuition, rather than a rigorous analysis of database variance. A study of key pharmacophore sizes for known ligand–protein complexes, plus an analysis of pharmacophore type distributions across our entire in-house database would likely produce more objective criteria for the filtering procedure. Nevertheless, the addition of filtering can clearly produce a further improvement in enrichment, although in this case to a lesser extent than normalization.

Conclusions

Binary pharmacophore fingerprints mask the highly uneven nature of pharmacophore distributions within the conformational ensemble of a molecule. Similarly many of the pharmacophores present in a fingerprint are common to most molecules or exist due to limitations in the pharmacophore construction and definition procedure. Consequently many of the pharmacophores added to a binary fingerprint do little to improve discrimination and likely add significant noise to any similarity calculation in which they take part. In the preliminary tests undertaken in this article, pharmacophore normalization and filtering are shown to mitigate this issue, producing significant improvements in lead hopping ability for the data set tested. While the resultant potential for such procedures is clearly highlighted, the tests presented here (in common with many articles on new descriptor technology) are too limited to provide deeper understanding with regard to descriptor performance, both from the perspective of virtual screening scenarios and alternative similarity

measures. For example, are there occasions when the procedure weights against infrequent pharmacophores that turn out to be bioactive? Does the inclusion of conformational flexibility data actually impact enrichment? With this in mind additional research has been undertaken to provide further insight into descriptor performance, highlighting a number of inherent issues with the current standards for descriptor validation. These efforts are presented in the following article.

References

1. Mason, J.S., Good, A.C. and Martin, E.J., Curr. Pharm. Des., 7 (2001) 567.
2. Good, A.C., Mason, J.S. and Pickett, S.D., In Böhm, H.-J. and Schneider, G. (Eds.), Virtual Screening for Bioactive Molecules, Wiley, New York, 2000, pp. 131–160.
3. Good, A.C. and Kuntz, I.D., J. Comput.-Aided Mol. Des., 9 (1995) 373.
4. Pickett, S.D., Mason, J.S. and McLay, I.M., J. Chem. Inf. Comput. Sci., 36 (1996) 1214.
5. Sheridan, R.P., Miller M.D., Underwood, D.J. and Kearsley, S.K., J. Chem. Inf. Comput. Sci., 36 (1996) 128.
6. McGregor, M.J. and Muskal, S.M., J. Chem. Inf. Comput. Sci., 39 (1999) 569.
7. Mason, J.S., Morize, I., Menard, P.R., Cheney, D.L., Hulme, C. and Labaudiniere, R.F., J. Med. Chem., 42 (1999) 3251.
8. Good, A.C. and Lewis, R.A., J. Med. Chem., 40 (1997) 3926.
9. Chem-Diverse, a module of Chem-X (version Jul-98B), part of the Accelrys software suite: www.accelrys.com.
10. Pozzan A., Feriani, F., Tedesco, G. and Capelli, A.M., In Höltje, H.D. and Sippl, W. (Eds.), Rational Approaches to Drug Design, Prous Science, Barcelona, Spain, 2001, pp. 224–228.
11. Concord 3D structure builder, distributed by Tripos and Optive Research Inc.: www.tripos.com, www.optive.com.
12. Pickett S.D., McLay, I.M. and Clark, D.E., J. Chem. Inf. Comput. Sci., 40 (2000) 263.