Similarity Searching in Files of Three-Dimensional Chemical Structures: Analysis of the BIOSTER Database Using Two-Dimensional Fingerprints and Molecular Field Descriptors

Ansgar Schuffenhauer, Valerie J. Gillet,* and Peter Willett

Krebs Institute for Biomolecular Research and Department of Information Studies, University of Sheffield,
Western Bank, Sheffield S10 2TN, U.K.

Received July 30, 1999

This paper compares the effectiveness of similarity measures based on two-dimensional fingerprints and on molecular fields for identifying pairs of bioisosteric molecules in the BIOSTER database. The results suggest that the two types of descriptor are complementary in nature, each finding some bioisosteric pairs that are not found by the other. This conclusion is confirmed by studies of groups of BIOSTER molecules that share the same activity characteristics, and by experiments that involve combining the two types of similarity measure.

INTRODUCTION

The calculation of intermolecular structural similarity lies at the heart of chemical similarity searching, where a target structure, typically a molecule with some biological activity, is searched against a database to find molecules that are structurally similar, and that are thus also expected to exhibit the activity of interest.[1,2] Many measures of molecular similarity have been discussed in the literature[3−6] but the most common measure, by far, involves the use of two-dimensional (2D) molecular fingerprints, bit strings that encode the presence of either predefined 2D substructures in a fragment dictionary or algorithmically specified chains of atoms. Although surprisingly effective in operation, such similarity measures are restricted in that many molecular properties, especially pharmacological ones, depend on the three-dimensional (3D) structure of the molecules. There has thus been much recent interest in the development of measures of 3D structural similarity.

Early 3D similarity measures were based on the matching of atoms using a range of geometric information (see, e.g., refs 7−10). Although simple in concept, such measures (like their simpler 2D counterparts) may miss molecules that exhibit similar biological effects, but that are based on different molecular skeletons; for example, the opioids have a rigid polycyclic structure but bind to the same receptors as the oligopeptidic natural endorphins. Examples of bioisosterism such as this suggest that a better measure of 3D similarity might be obtained by considering the electrostatic, steric, and hydrophobic molecular fields that form the principal inputs to modern methods for 3D QSAR,[11] rather than the geometric arrangement of the atoms comprising a molecule. Following the pioneering work of Carbó and his collaborators,[12] there has been much interest in techniques (see, e.g., refs 13−17) that involve aligning the target structure with a database structure to find the superposition that maximizes the value of the chosen similarity measure. This alignment procedure is repeated for each of the database structures in turn, so as to find those that have the most similar 3D pattern of field descriptors. The identification of the best possible alignment involves a search of translational and rotational space (and also torsional rotation space if conformational flexibility is to be included in the similarity measure), and is thus extremely demanding of computational resources unless an efficient search algorithm can be identified.

Work in Sheffield has resulted in the development of a program called FBSS (for field-based similarity searching) that uses a genetic algorithm (GA) for the alignment of molecular fields.[18−20] A GA is a computational procedure that encodes potential solutions to the problem that is to be optimized in data structures called chromosomes, and then processes populations of such chromosomes by means of mutation and crossover operators analogous to those encountered in Darwinian evolution.[21−24] Each chromosome has an associated fitness score, which represents the extent to which that particular chromosome provides a solution to the problem that is to be investigated, and application of the genetic operators over many generations results in an increase in the average fitness of the members of the population. Convergence of the fitness scores to a maximum value indicates the identification of the best solution to the problem under investigation. In FBSS, a chromosome encodes a potential alignment of the target structure and a database structure by means of the translations and rotations that would be required to fit one to the other, and the fitness function is the Carbó similarity index for the overlap of the fields (electrostatic, steric, or hydrophobic fields or any combination thereof) resulting from the encoded alignment.[20] The Carbó index, S_C, is a form of the long-established cosine coefficient, and is given by the equation

$$S_c = \frac{\int P_A(r)\, P_B(r)\, dr}{\int P_A(r)\, P_A(r)\, dr^{1/2} \int P_B(r)\, P_B(r)\, dr^{1/2}}$$

* To whom correspondence should be addressed. E-mail: v.gillet@sheffield.ac.uk.

10.1021/ci990263g CCC: $19.00
Published on Web 12/22/1999

where P_A and P_B are the properties of the two molecules, A and B, that are being compared, for example electrostatic potential. The potentials are calculated using the Gaussian approximation approach described by Good et al.[14]

Searches of the World Drugs Index (WDI) database[25] using both FBSS and the 2D fingerprints in the UNITY chemical information management system produced by Tripos Inc.[26] suggested that the latter similarity measure tended to find more bioactive molecules, but that those resulting from the field-based measure were structurally more diverse. In this paper, we report a further comparison of field-based and 2D-based similarity searching, focusing here upon the ability of the two measures to identify molecules that are known to exhibit bioisosterism.[27, 28]

EXPERIMENTAL DETAILS

Our previous studies have considered just a small number of target structures that were searched against the WDI database, which contains a wide range of types of molecule that have been reported in the literature as exhibiting biological activity. The experiments here used a very different type of file, specifically the collection of literature references that has been gathered by Prof. István Ujvary over the past 35 years[29] and that is available as the BIOSTER database from Synopsys Scientific Systems.[30] BIOSTER contains 2D structures for ca. 5000 molecules together with ca. 3500 bioisosteric relationships between pairs of these molecules (with the relationships recorded in a statement from the literature that one molecule is a bioisostere of another in the database), and the database hence presents an ideal source of data for evaluating computer methods for identifying bioisosteres. The database was processed as two MDL RD-File[31] files. The first contains each molecule as a 2D structure diagram with chirality descriptors and a molecular identification (ID), and the second represents each bioisosteric relationship as if it was a reaction, with the IDs of the “reactant” and “product” molecules, the literature reference, and keywords describing the biological function of the pair of molecules.

The molecules from the first of the BIOSTER files were converted to a SYBYL database, with the 3D structures being generated using the CONCORD program[26] followed by optimization with the PM3 routines in MOPAC.[32] Some experiments were also carried out using SYBYL MAXMIN and Gasteiger point charges, but both methods gave analogous results so we consider here just those obtained in the PM3-based experiments; in similar vein, we have not reported results obtained from the use of CORINA,[33] rather than CONCORD, for the generation of the 3D structures. Some molecules could not be processed (e.g., there was an unspecified residue in a molecule, nonstandard elements were present, a molecule contained two or more nontrivial structural fragments, or the molecule was a radical) and these were discarded, yielding a set of 5023 molecules and 2995 bioisosteric relationships (i.e., “reaction” pairs) for the experiments.

The similarities between pairs of molecules in a “reaction” were then calculated using UNITY’s conventional 2D fingerprint similarity measure and using the field-based similarity measures in FBSS. Four types of FBSS measure were employed—electrostatic, steric, hydrophobic, and all three combined (referred to subsequently as “All”)—with the GA that lies at the heart of FBSS being parametrized with a population of 125 chromosomes, 10 000 operations, and a selection pressure of 1.1.[20]

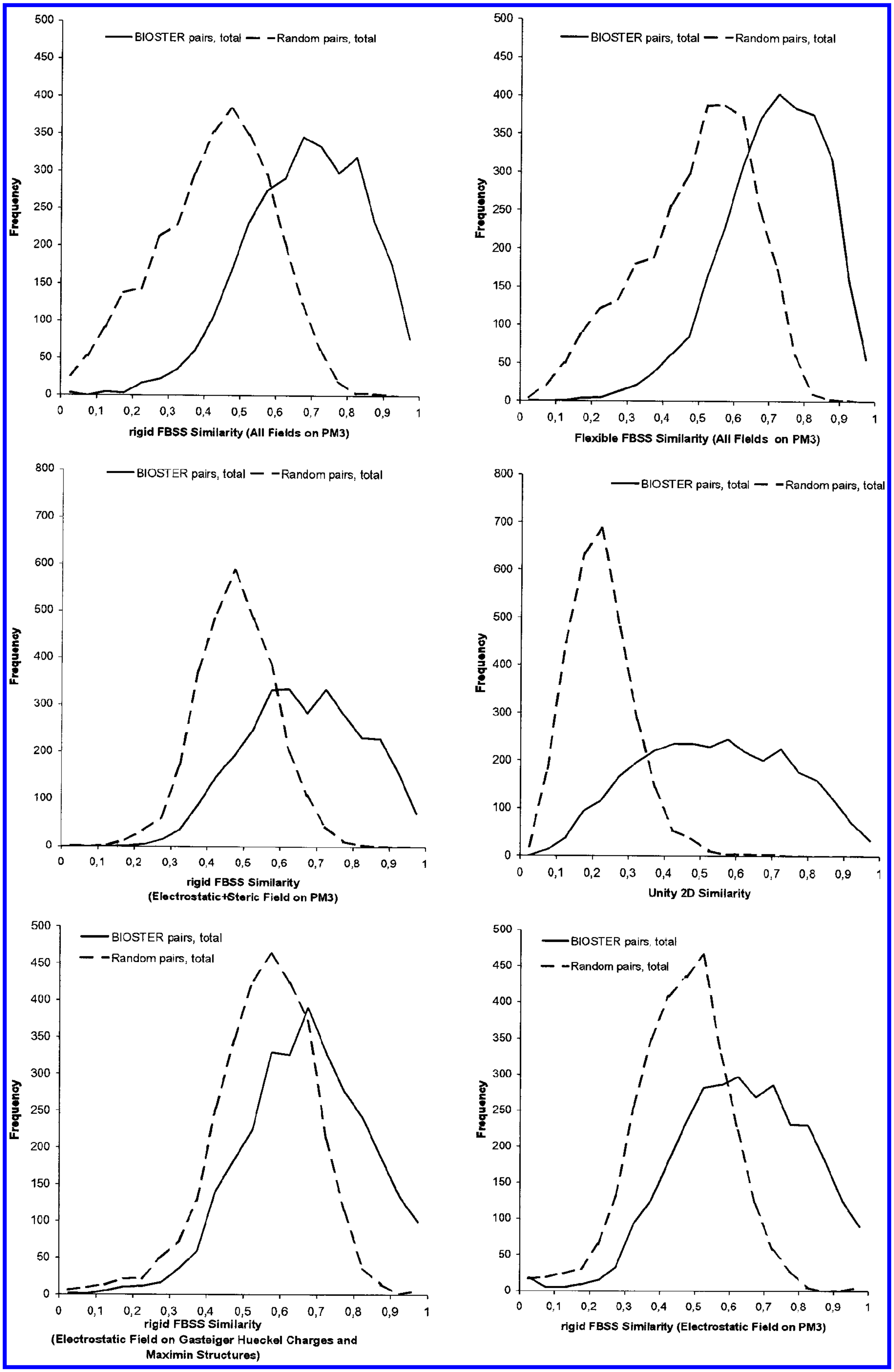
Table 1. Mean Similarities and Standard Deviations Averaged over 2995 Pairs of BIOSTER Molecules (either a Bioisosteric Pair or a Pair Selected at Random)

	$\mu \pm \sigma$	
similarity measure	BIOSTER pairs	random pairs
UNITY	0.54 ± 0.21	0.22 ± 0.09
electrostatic	0.63 ± 0.19	0.46 ± 0.14
steric	0.77 ± 0.12	0.56 ± 0.29
hydrophobic	0.82 ± 0.18	0.65 ± 0.10
electrostatic/steric	0.67 ± 0.16	0.48 ± 0.11
all	0.68 ± 0.16	0.43 ± 0.16
all (flexible)	0.71 ± 0.14	0.50 ± 0.16

INITIAL COMPARISON OF 2D AND FBSS SIMILARITIES

Our initial experiments calculated the UNITY and FBSS similarities for the set of 2995 reaction pairs. For comparison, the resulting similarities were compared with those calculated for 2995 pairs of molecules selected at random from BIOSTER, subject only to the condition that they were not linked by a “reaction”. The results are summarized in Table 1, which lists the mean (μ) and standard deviation (σ) for the BIOSTER and random distributions, averaged over the 2995 similarities in each case. The UNITY similarities are calculated using the Tanimoto coefficient and fall in the range 0−1. The FBSS similarities are calculated using the Carbó coefficient, and although negative values are theoretically possible they rarely occur because of the alignment procedure in FBSS which will tend to spatially separate any pair of molecules for which a positive similarity score is not possible. The result of this will be a near-zero similarity score, and any negative values are not shown in the plots in Table 1. It will be seen that the BIOSTER similarities are consistently and significantly ($p < 0.001$ in a Z test for the difference of two means) greater than the random similarities for all of the measures considered here. This conclusion is supported by the frequency distributions shown in Figure 1, each part of which shows the distribution of similarities for the 2995 bioisosteric pairs (the solid line) and for the 2995 randomly selected pairs (the dashed line). It will be seen that the solid lines are consistently to the right of the dashed lines, and we may hence conclude that the UNITY and FBSS measures generally identify the bioisosteric pairs as being more similar to each other than are the random pairs.

Having established that both UNITY and FBSS perform better than random, it is, however, also clear that there is a fair degree of overlap between the pairs of distributions. This effect is most obvious in the case of the steric field measure where there is no part of the distribution that corresponds to a complete separation of BIOSTER and random pairs. For all of the other measures, conversely, there is at least some region of the distributions where there is such a complete separation. One explanation for the different characteristics of the distribution derived from the steric fields is the fact that steric fields are measured on a scale of 0 to 1, whereas the electrostatic and hydrophobic fields are measured on the scale −1 to +1, and no explicit penalty is made in the Carbó


BIOSTER pairs, total
Random pairs, total
Frequency
0
50
100
150
200
250
300
350
400
450
500
0 0,1 0,2 0,3 0,4 0,5 0,6 0,7 0,8 0,9 1
rigid FBSS Similarity (All Fields on PM3)
BIOSTER pairs, total
Random pairs, total
Frequency
Flexible FBSS Similarity (All Fields on PM3)
BIOSTER pairs, total
Random pairs, total
Frequency
0
100
200
300
400
500
600
700
800
rigid FBSS Similarity
(Electrostatic+Steric Field on PM3)
BIOSTER pairs, total
Random pairs, total
Frequency
Unity 2D Similarity
BIOSTER pairs, total
Random pairs, total
Frequency
rigid FBSS Similarity
(Electrostatic Field on Gasteiger Hueckel Charges and Maximin Structures)
BIOSTER pairs, total
Random pairs, total
Frequency
rigid FBSS Similarity (Electrostatic Field on PM3)

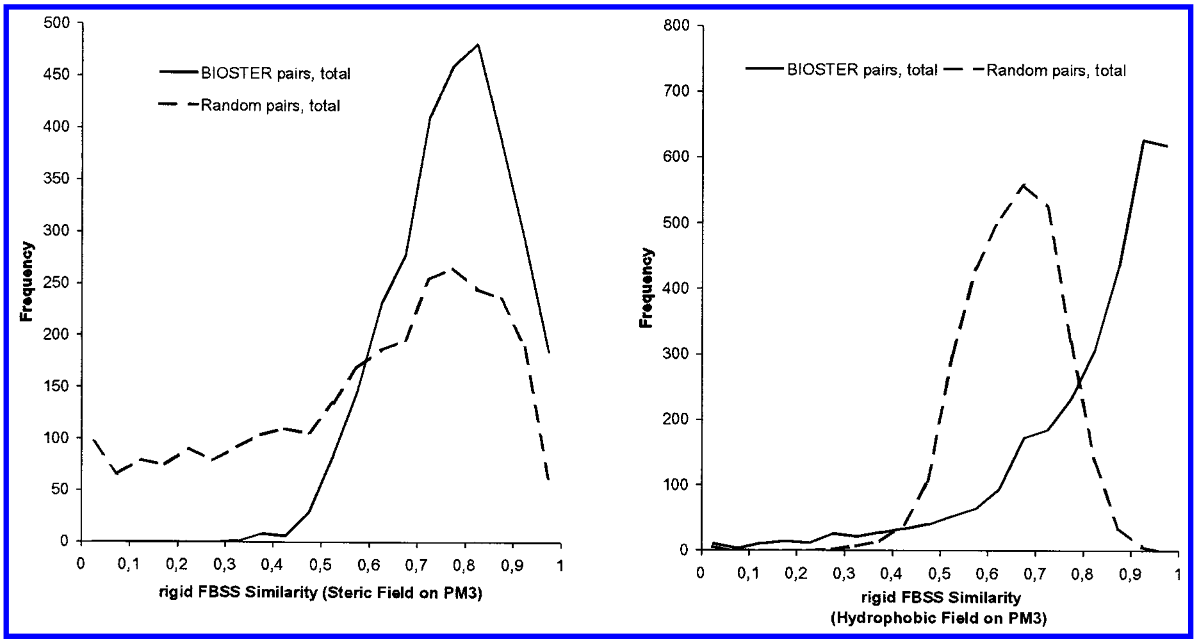


Figure 1. Frequency distributions for the similarities of pairs of BIOSTER molecules: solid plots are for bioisosteric pairs and dashed lines for pairs selected from the database at random.

Table 2. FBSS Similarities Averaged over 54 Bioisosteric Pairs, for Experimental (CSD) and Calculated (CONCORD/PM3) Geometries

geometry	electrostatic	steric	hydrophobic	all
CSD	0.62 ± 0.24	0.79 ± 0.23	0.76 ± 0.10	0.64 ± 0.19
CONCORD/PM3	0.72 ± 0.17	0.82 ± 0.22	0.86 ± 0.08	0.73 ± 0.18

coefficient when one part of a molecule occupies space that is not occupied by the other. Conversely, for electrostatic fields, the overlap of a positively charged area of one molecule with a negatively charged part of another molecule is penalized explicitly. A similar effect occurs for the hydrophobic fields, where the overlap of a hydrophobic region in one molecule with a hydrophilic region in the other is penalized explicitly.

In the case of the hydrophobic measure, the region of separation of the FBSS and the UNITY distributions is extremely small (0.9−1) and also corresponds to the region where the majority of the BIOSTER-pair similarities occur. The small range of values found for the FBSS similarities suggests that the hydrophobic measure is less suitable for database searching, where compounds are ranked in decreasing order of similarity, than the other FBSS measures. This is because even very slight changes in the magnitude of a similarity (arising, e.g., from an imperfect alignment) can cause major changes in the rankings. This conclusion confirms the observations of Drayton et al.[20] that the hydrophobic field was the least discriminating of those tested.

In the case of the UNITY measure, the region of separation is very marked, and this might be taken to suggest that this 2D similarity measure is best able to identify bioisosteres from among the measures considered here. One explanation for this apparent superiority is that BIOSTER contains many pairs of molecules that are topologically similar as a result of the common phenomenon of the so-called "me-too" or "fast-follower" drugs and 2D similarity measures are ideally suited to finding these topological similarities. Another reason for the apparent poorer performance of the field-based methods is that they require much more precise structural data for their evaluation, in the form of a 3D conformation, than the 2D structure diagram required by UNITY and similar measures. A further limitation of the electrostatic and All versions of FBSS is that the existence of a charge in a molecule will tend to swamp any other structural (dis)-similarities; related to this is the protonation status of a molecule, which will depend on its environment but which is generally ignored in BIOSTER, where molecules are stored in their neutral form. Thus, if one carries out a database search for acetylcholine, then many uncharged amines that are bioisosteres are missed; instead, the nearest neighbors are typically other molecules with a net positive charge, irrespective of any other structural features that they may possess.

Given the dependence of rigid FBSS on 3D conformation, it might have been expected that FBSS would perform better the more accurate the conformations that were employed. As noted previously, the 3D structures were generated using CONCORD; however, an alternative source of 3D coordinates is the Cambridge Structural Database (CSD)[34] of X-ray crystallographic structure determinations. A total of 54 pairs of molecules from BIOSTER was found where both molecules also occurred in the CSD, and we hence calculated similarities for these bioisosteric pairs using the CSD structures and compared the results with the similarities calculated using CONCORD/PM3. It was found, as shown in Table 2, that the similarity values obtained from the experimental geometries were smaller than those based on the calculated geometries. From this we conclude that the greater consistency resulting from calculated geometries (where the same substructural moieties will always be represented in exactly the same way) are preferred to the experimental geometries, where the structures are obtained under varying conditions (e.g., solvent or pH). Note that we should ideally employ the bound conformation of the ligand, but this is, of course, not generally available at the stage of a lead-discovery program where similarity-based methods (such as those considered here) are called upon to play a role. Experiments were also carried out using the flexible fitting routine in FBSS,[19] which allows for torsional rotations and which will thus permit exploration of the bound conformation (whatever that might be); however, this was not found to result in any increase in performance, as shown by the bottom row of Table 1. In part, this can be explained by known limitations in the FBSS routine, which does not include ring-flipping and which has just a simple bump-check to constrain the generation of high-energy conformations. In addition, we have found that the GA requires large populations and many generations if it is to give reasonable results with flexible structures (presumably because of the very much larger search space that needs to be explored), and the resulting run times make it infeasible for use in a database-searching context.

There is a further characteristic that affects the relative performance of the FBSS and 2D similarity measures, viz., the sizes of the molecules that are being compared. Inspection of the BIOSTER pairs that FBSS found to be highly similar to each other showed that they mostly involved relatively small molecules, whereas highly dissimilar pairs tended to involve large, flexible molecules. The BIOSTER and random pairs were hence partitioned into a set of bins, depending on the molecular size (specifically, the number of atoms including hydrogens) of each of the partners. The results of this partitioning are shown in Figures 2−4, where it will be seen that the 2D and FBSS similarity measures show a different size dependency.

The UNITY 2D similarities increase with molecular size for both the random pairs and the BIOSTER pairs, owing to the increasing saturation of the bit strings that has been discussed in detail by Flower.[35] The FBSS similarities for the BIOSTER pairs decrease rather slowly as the molecular size increases, up to about 50 atoms (including hydrogens), but faster thereafter. There are at least three possible reasons for this behavior. First, and most obviously, the property under consideration may be size-dependent. This is certainly the case with the hydrophobic field, where small molecules are mostly hydrophilic but where larger molecules exhibit a wider range of hydrophobicities. Thus smaller molecules will tend to have higher average hydrophobic similarities than larger molecules, as illustrated in Figure 3. Second, the greater the structural complexity and the greater the number of rotatable bonds, the more difficult it is for structure generators such as CONCORD to produce high-quality conformations for large molecules. Finally, we have found that the ability of FBSS to align molecular fields decreases with an increase in molecular size, as demonstrated in Table 3. Here, 16 molecules were selected at random from the BIOSTER database in each of the following size ranges: ≤20, 21−30, 31−40, 41−50, 51−60, 61−80, and >80 atoms (again including hydrogens). FBSS then tried to align each molecule with itself, and the variation in successful self-alignment with molecular size was noted, where a successful self-alignment was judged to be a similarity >0.97 obtained using the standard GA parameters. Table 3 shows clearly the decreasing effectiveness of the alignment procedure as larger and larger molecules are involved, especially when the flexible version of the GA is used (as noted previously by Thorner et al.[19]).

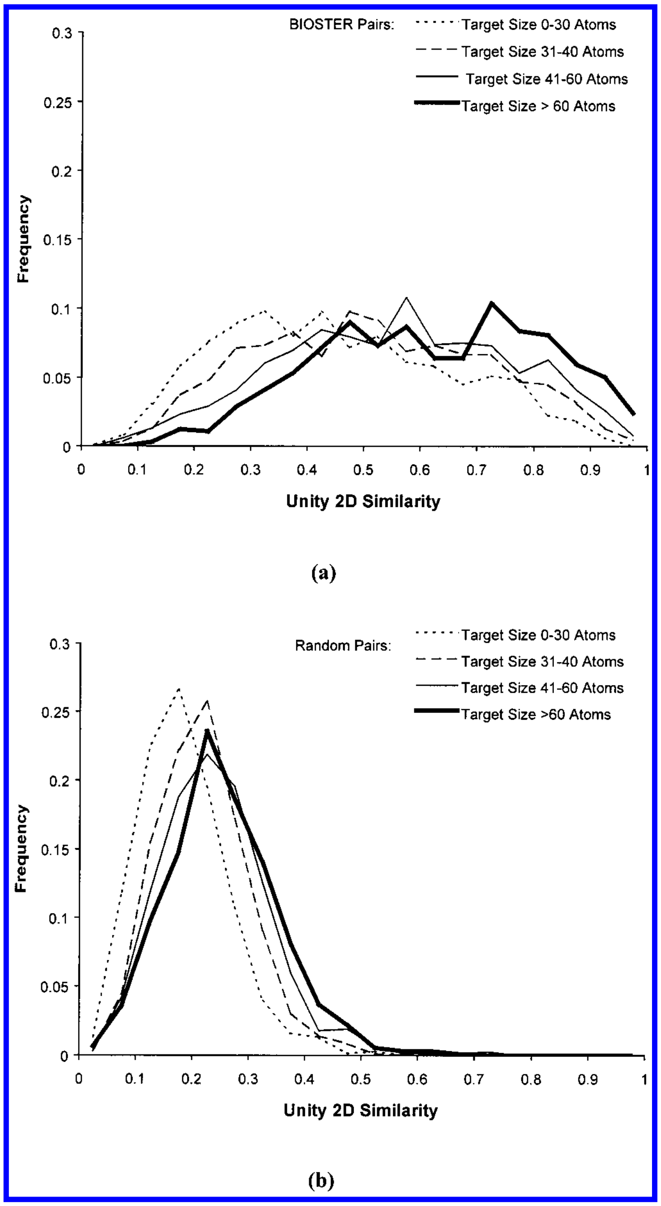
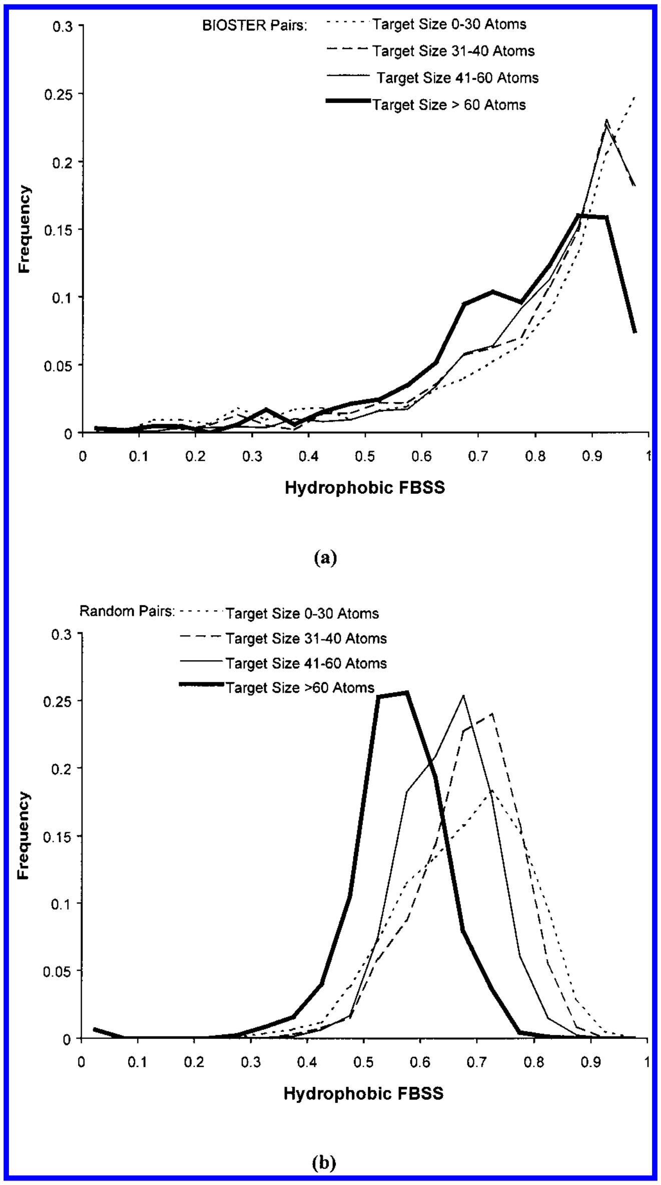


Figure 2. Frequency distributions for the UNITY similarity measure for different molecular sizes within (a) BIOSTER pairs and (b) random pairs.

Table 1 shows the performance of the 2D and FBSS measures when averaged over all 2995 pairs of molecules; however, there are also substantial differences between the two types of measure when one considers their performance on individual pairs of molecules. Figure 5 is a scatter plot for the 2D and FBSS similarities. The correlation coefficient for the plot is $r^2 < 0.1$, and thus while there is some degree of correlation between the two sets of similarity values, there are a fair number of cases where a high 2D similarity is associated with a low FBSS similarity, and vice versa. Such cases would suggest some degree of complementarity between the two types of similarity measure: this possibility is considered further below.

ANALYSIS OF ACTIVITY CLASSES

As noted previously, the BIOSTER database is organized around pairs of molecules that have been reported as having



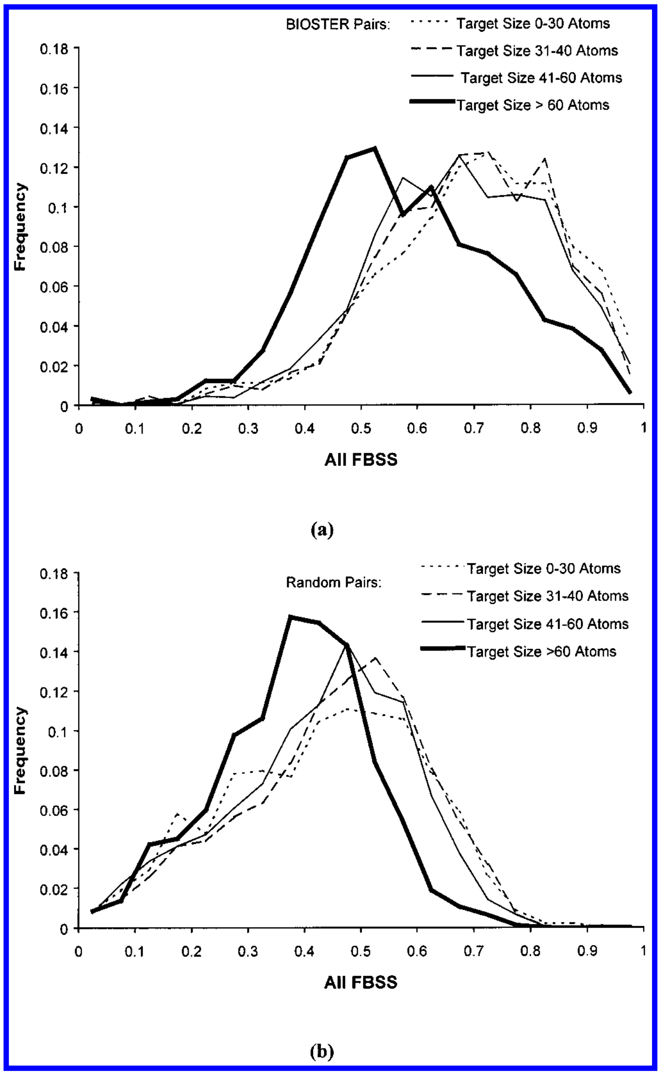
(a)

(b)

Figure 3. Frequency distributions for the hydrophobic FBSS similarity measure for different molecular sizes within (a) BIOSTER pairs and (b) random pairs.

the same type of activity. However, given a total of, say, N molecules that have been assigned the same activity, there are normally many less than $N(N-1)/2$ pairs of molecules recorded in the database as being bioisosteric. The bioisosteric relationships that have been identified can arise in many ways,[27,28] including "me-too" compounds, for example, new leads that are related to an existing drug; synthetic compounds that are related to the natural ligand of a receptor or the natural substrate of an enzyme; and compounds that are the result of structural variations in parts of the molecules that are of less importance for receptor binding but that have been introduced to improve the pharmacokinetic behavior of a molecule. There are also a small number of relationships identified in BIOSTER that are not bioisosteric relationships as such, for example, a molecule that is related in some way to its prodrug or tautomer.

If the pairs of molecules recorded in the database for some specific activity class arise principally from a "me-too" relationship, then the average similarity of the BIOSTER pairs belonging to this class may be different from the average similarity for all pairs of molecules within that class; in particular, the close topological relationships between "me-too" compounds might be expected to inflate the average 2D similarities (but there might also be concomitant effects for the FBSS similarities). We have thus tried to overcome this potential problem by grouping all molecules that have some specific activity, as denoted by their keywords. The level of description within BIOSTER varies considerably, including both macroscopic, pharmacological effects (e.g., "antibiotic") and detailed mechanistic classes (e.g., "alanine-alanine transpeptidase inhibitor"). We have focused on the latter, more detailed type of description and selected nine classes for examination where there are sufficient numbers of molecules to enable substantive conclusions to be drawn. Each compound in a class was compared with every other compound in the class using both the UNITY and FBSS



(a)

(b)

Figure 4. Frequency distributions for the All FBSS similarity measure for different molecular sizes within (a) BIOSTER pairs and (b) random pairs.

Table 3. Number of Structures Failing the Self-Alignment Test (Similarity < 0.97) Is Dependent on Molecular Size (16 Molecules per Size)

	size (including H)						
	≤20	21−30	31−40	41−50	51−60	61−80	>80
all (rigid)	0	2	0	1	5	7	5
all (flexible)	4	7	11	12	14	16	16

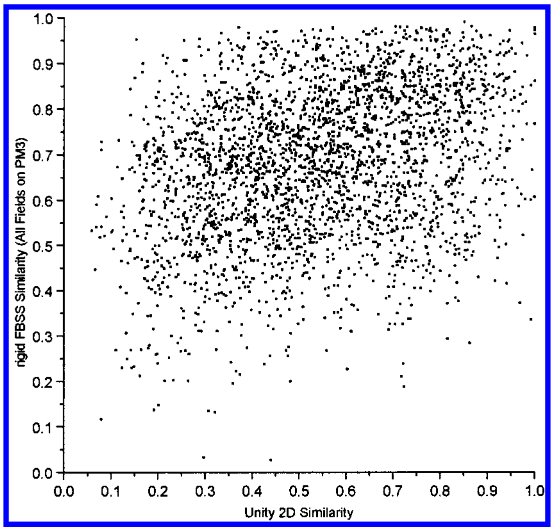


Figure 5. FBSS similarities (all fields, rigid) vs UNITY similarities for the BIOSTER pairs.

similarities, with the resulting average values listed in Table 4. It will be seen that the UNITY values within each specific class are generally significantly less than when all 2995 BIOSTER pairs are considered and there is also a reduction, but a noticeably smaller one, in the FBSS similarities, thus confirming the high incidence of "me-too" relationships in the BIOSTER pairs. We now consider the various activity classes in more detail.

Adrenaline Analogues (Figure 6). This activity class contains agonists and inhibitors of adrenergic receptors, and in some cases the molecules are noted as being selective for either the α- or *β*-receptor. Most of the molecules contain a benzene ring and an amino group, as found in noradrenaline (**1**), and in the case of α-receptor binding compounds the amino group may be incorporated in a ring system (**5**) or replaced by a imino group (**6**). The FBSS similarities and UNITY similarities of these compounds were found to be average when compared with other activity classes. Neither FBSS nor UNITY were able to discriminate between agonists and inhibitors or between α- and *β*-selective compounds, except that the average UNITY similarities among the *β*-receptor selective compounds were significantly higher than the other members of the class. This is probably due to the fact that the *β*-agonists and inhibitors are structurally less diverse than the other adrenergic compounds and they often share at least the *tert*-butyl group as a common structural feature.

Bacterial Alanine-Alanine Transpeptidase Inhibitors (Figure 7). The best known representative of this class of drugs is penicillin (**9**). All of these compounds share the *β*-lactam ring system as a structural feature and thus show rather high UNITY similarity, compared to the other classes. Also, there are examples of structural modifications present in this class, which are mainly introduced to improve the pharmacokinetic properties rather than the binding to the transpeptidase itself (e.g., **10**). Such modifications do not affect the recognition of the key functionality in the fingerprint, but they do make the 3D alignment of the molecules more difficult. The similarities were also calculated between the antibiotics in the class and a fragment of the natural product of the alanine-alanine transpeptidase (**12**) which does not contain the *β*-lactam ring. Whereas the average similarity calculated by FBSS is in the same range as the average similarity between the antibiotics, the average UNITY similarity between the dipeptide and the antibiotics is lower than that between the two antibiotics. This shows that the higher average UNITY similarity for the antibiotics is mainly based on the characteristic *β*-lactam ring as a common structural fragment.

Bacterial Dihydropteroidin Reductase Inhibitors (Figure 8). The compounds in this class are used as antibiotics and are, with the exception of (**23**), *p*-aminobenzenesulfonic acid amides. Most of the structures have an aromatic ring system that is connected to the amide nitrogen (e.g., **24** and **25**), and most structural variations involve replacement of this ring system with another heteroaromatic ring. The UNITY similarities are rather low despite the fact that most structures share a common 2D feature in the form of the sulfonamide group: this can be explained by the fact that the heteroatom substitutions cause major changes in molecular fingerprints. FBSS is less sensitive to this effect since it is concerned with the fields generated by the atoms; it additionally benefits from the rigidity of the aromatic ring systems, which facilitates the alignment with the result that the calculated similarities are high. FBSS is also able to recognize the similarity between the natural substrate of the enzyme, *p*-aminobenzoic acid (**22**) and the sulfonamides, whereas the UNITY similarities are no better than random.

DNA Intercalators. This class contains molecules that can bind between two layers of base pairs in DNA and that are used as anticancer drugs. All substances contain a large polycyclic aromatic system, but other parts of the structures are highly variable. The FBSS All similarities and the UNITY similarities were both no better than random. The FBSS steric similarities were better than random, but this is probably not significant owing to the high scattering of the random values for this descriptor. However, the ability to intercalate in DNA is dependent on the presence of the large flat ring system, which should lead to a similar shape in at least a major part of the molecules.

Enkephalins and Opioids (Figure 9). This class contains the opioid analgesics and fragments of the natural agonists at their receptors, the enkephalins (Leu-enkephalin (**26**)), and their structural variations (e.g., **27**). Enkephalins are pentapeptides which always contain a tyrosine at the amino terminus and are in vivo active drugs; many of the opioids contain a phenolic OH or OCH_3 group (e.g., **28**). At the 2D level the structural discrepancies between the enkephalins on one hand and the opioids on the other are large, and the UNITY similarities for the class are only average. The average UNITY similarity among the peptides is high due to their similar building blocks, but the highly flexible nature of the peptides means that FBSS was not expected to perform well (since it is dependent on 3D conformation). This was indeed found to be the case with the average FBSS similarity among the peptides being no better than random. In the nonpeptidic subclass, however, FBSS performed well, since it certainly benefits from the rigidity of the polycyclic morphine analogues, although many synthetic opioids (e.g., **29**) have a less rigid structure than morphine.

Estrogens (Figure 10). The molecules in this activity class are characterized by a rigid skeleton and a low number of functional groups, and the former characteristic results in higher than average FBSS similarities. The steroid ring

Table 4. Results for Selected Activity Classes

activity class (number of compounds)	number of pairs	rigid FBSS all field similarity	UNITY similarity
all BIOSTER pairs	2995	0.68 ± 0.16	0.54 ± 0.21
all random pairs	2995	0.43 ± 0.16	0.22 ± 0.09
adrenaline analogues (58) total	1653	0.57 ± 0.12	0.34 ± 0.13
agonist ↔ agonist	465	0.57 ± 0.12	0.36 ± 0.14
inhibitor ↔ inhibitor	351	0.57 ± 0.11	0.35 ± 0.14
agonist ↔ inhibitor	837	0.56 ± 0.11	0.32 ± 0.12
α ↔ β	91	0.60 ± 0.12	0.27 ± 0.17
β ↔ β	153	0.54 ± 0.10	0.47 ± 0.10
α ↔ β	252	0.54 ± 0.10	0.27 ± 0.10
bact. alanine-alanine transpeptidase inhibitors (74)	2701		
inhibitor ↔ inhibitor	2701	0.50 ± 0.11	0.44 ± 0.14
nat. substrate ↔ inhibitor	74	0.53 ± 0.09	0.29 ± 0.07
bact. dihydropteroidin reductase inhibitors (14)			
inhibitor ↔ inhibitor	91	0.66 ± 0.15	0.36 ± 0.17
nat. substrate ↔ inhibitor	15	0.58 ± 0.23	0.24 ± 0.13
DNA Intercalators (10)	45	0.37 ± 0.20	0.23 ± 0.18
		0.64 ± 0.14[a]	
enkephalins/opioids (42) total	861	0.46 ± 0.15	0.45 ± 0.14
nonpeptide[b] ↔ nonpeptide	276	0.61 ± 0.09	0.43 ± 0.13
peptide ↔ peptide	153	0.47 ± 0.12	0.63 ± 0.15
peptide ↔ nonpeptide	432	0.36 ± 0.11	0.39 ± 0.08
estrogens (34) total	561	0.72 ± 0.10	0.45 ± 0.17
agonist ↔ agonist	210	0.76 ± 0.11	0.49 ± 0.19
inhibitor ↔ inhibitor	58	0.69 ± 0.09	0.42 ± 0.15
agonist ↔ inhibitor	228	0.72 ± 0.10	0.43 ± 0.15
glucosidase inhibitors (18)	153	0.65 ± 0.11	0.30 ± 0.15
		0.63 ± 0.11[c]	
		0.58 ± 0.18[d]	
HMG-CoA reductase inhibitors (10)			
active structures[e]	45	0.65 ± 0.07	0.37 ± 0.17
		0.62 ± 0.07	0.34 ± 0.15
NMDA antagonists (44)	946	0.51 ± 0.17	0.26 ± 0.11
glycine site ↔ glycine site	325	0.58 ± 0.14	0.30 ± 0.13
glutamate site ↔ glutamate site	66	0.58 ± 0.12	0.36 ± 0.05
glutamate site ↔ glycine site	240	0.47 ± 0.14	0.22 ± 0.07

[a] FBSS similarity using steric fields only. [b] "Peptide" here includes close structural analogues as well as the oligopeptides. [c] Structures manually aligned including the modification of torsion angles. [d] Structures manually aligned without modification of torsion angles. [e] The lactone ring was opened and the structures were reoptimized (since one structure was present both as lactone and open ester, there is one structure less in the active set).

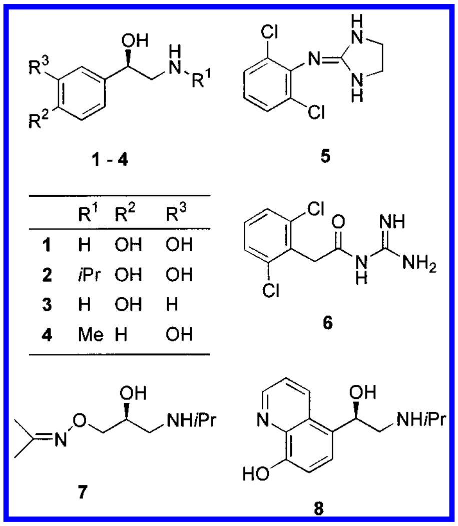


Figure 6. Adrenaline analogues.

system that is present in many of the molecules (e.g., in **30**) enhances the recognition by 2D similarity, and even structures without the steroid ring system (e.g., **31**) are recognized by UNITY, since the modified skeletons still have aromatic and aliphatic carbon−carbon bonds in a similar, low functionalized chemical environment leading to similar connection paths. Neither FBSS nor UNITY is able to distinguish between active compounds and inhibitors.

Glucosidase Inhibitors (Figure 11). The six-membered ring of glucose (**38**) with its characteristic substitution pattern is at least partially present in all the structures in this class. The most common variations are the substitution of the endocyclic oxygen by another heteroatom (e.g., **39**) or the substitution of the endocyclic C−O bond by an imine (e.g., **40**) to imitate the transition structure during the breaking of the glycosidic bond. The average UNITY similarity is rather low due to the heteroatomic substitutions, whereas the rigidity of the ring system again leads to high FBSS similarities. Since the structural analogy in this class is rather high, it was possible to try a manual alignment of the structures; however, this resulted in slightly lower field-based similarities than the alignment found using rigid FBSS. In fact, even after modifying the remaining torsions no similarity higher than the rigid FBSS similarity could be achieved.

HMG-CoA Reductase Inhibitors (Figure 12). HMG-CoA reductase is one of the enzymes involved in the biosynthesis of cholesterol. It is inhibited by cholesterol itself, and so some close analogues of cholesterol are included (e.g., **43**) in this class of molecules. The other molecules are either derivatives of natural products such as lovastatine (**41**) or totally synthetic aromatic structures (**44**). Most of the inhibitors contain a 3,5-dihydroxycarboxylic acid as side chain which exists in its open chain form in the active compounds. However, many of the inhibitors are stored in

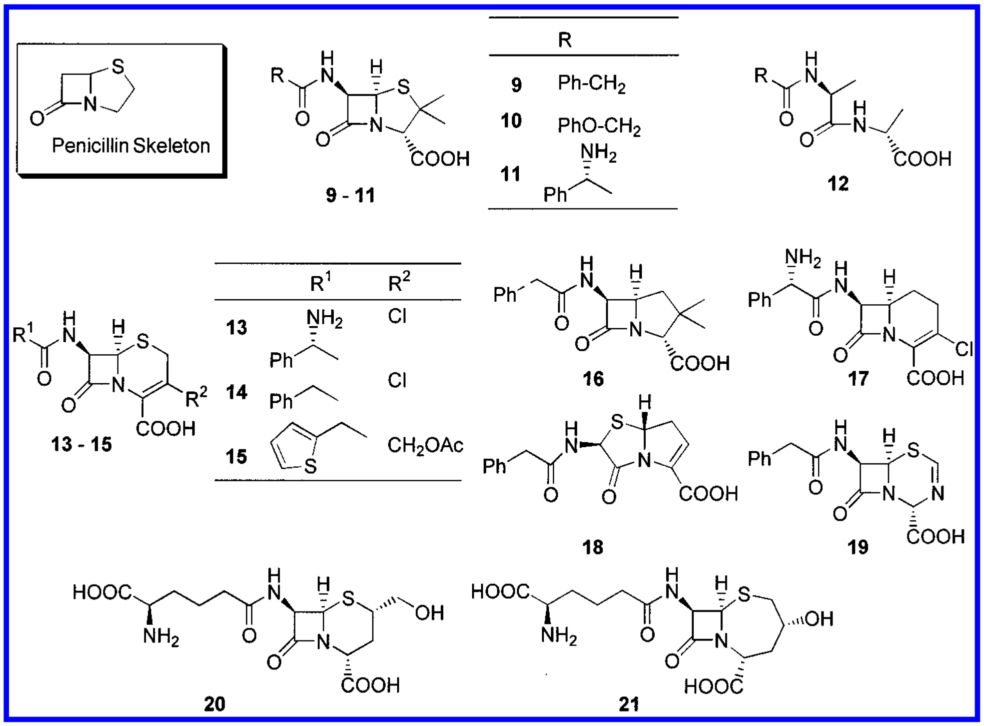


Figure 7. Inhibitors of bacterial alanine-alanine transpeptidase (β-lactam antibiotics).

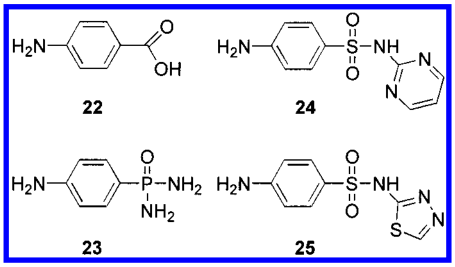


Figure 8. Inhibitors of bacterial dihydropteroidin reductase.

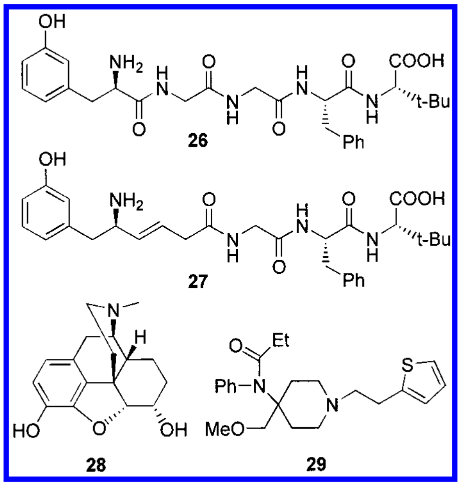


Figure 9. Enkephalins and opioids.

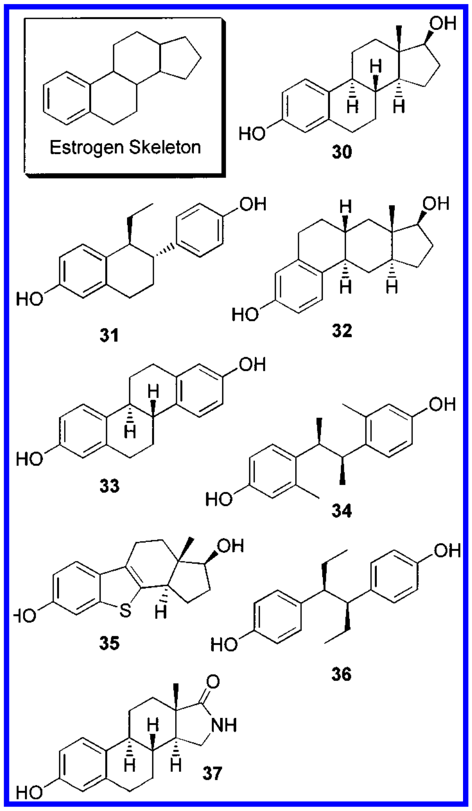


Figure 10. Estrogens.

BIOSTER as lactones, which are transformed in vivo into the active form (**41** → **42**). Half of the molecules, therefore, are stored in BIOSTER as lactones. Although the FBSS similarities were quite high when calculated using the molecules as found, the different forms that are possible might reduce the similarity. Therefore, open chain structures were generated for all the molecules. This, however, did not improve the FBSS similarities, from which we conclude that what might have been gained by removing a source of structural diversity might then have been lost by transforming a rigid cyclic fragment into a more flexible open chain.

NMDA (*N*-Methyl-D-aspartate) Antagonists. These substances are used to prevent neuronal degeneration by blocking the NMDA receptor. The UNITY similarities are no better than random and the FBSS similarities are only moderate, when compared to other classes. Although the structures are quite diverse, most of them are not very flexible so that a high average FBSS similarity might have been expected. Since a wider scattering of the FBSS similarity

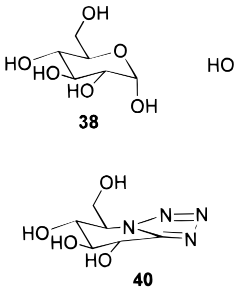


Figure 11. Glucosidase inhibitors.

values was observed than for any other class, there is a possibility that the compounds may actually belong to more than one distinct class. An examination of the references given in BIOSTER revealed that the compounds do bind to different sites of the NMDA receptor, which has a glutamate and a glycine binding site,[36] as well as other binding sites for noncompetitive antagonists.[37] Once the compounds were grouped according to their binding sites (glycine and glutamate—all others were omitted as their binding sites were often only vaguely specified in the literature references), both the FBSS and UNITY similarities improved. The cross-similarity was no better than random, which was expected, since in contrast to the subtypes of the adrenaline receptors, two completely different binding sites were involved.

COMBINATION OF UNITY AND FBSS SIMILARITIES USING DATA FUSION

Data fusion[38,39] is the process of combining different sources of information into a single decision function that can provide better estimates than the individual sources themselves. The technique has been applied in the context of chemical similarity by combining the rankings resulting from use of different similarity measures in a chemical similarity search.[40] Such combined rankings have been found to exhibit a level of performance that is, on average, superior to that of any individual similarity measure.[41] Here, we consider the fusion of FBSS and UNITY similarities; the first experiments we performed fused the actual similarity scores output by each of the measures, and we then compared these results with fusion of the rankings produced by each similarity measure.

An inspection of Table 1 shows that UNITY similarities are generally less than the corresponding FBSS similarities, and we must hence find some way of scaling them prior to data fusion if both types of measure are to contribute equally to the final combined score. The UNITY similarity values cover the full range from 0 to 1, but the FBSS similarities are much less evenly distributed (especially so in the case of the hydrophobic similarities where most of the values are concentrated in the upper half of the possible range of values, as shown in Figure 1). A transformation is thus required so that the two sets of similarities cover comparable ranges of values; specifically, the transformation must compress the lower range of FBSS similarity values while expanding the higher range and we have chosen to use a quadratic function for this purpose. A further justification for using such a function arises from a consideration of the natures of the coefficients that are used to calculate the two measures, viz., the Tanimoto coefficient for the UNITY measure and the Carbó coefficient for the FBSS measure.

The simplest forms of the coefficients, as applicable to bit strings, are

$$S_{\text{cosine}} = \frac{C}{\sqrt{AB}} \quad \text{and} \quad S_{\text{Tanimoto}} = \frac{C}{A + B - C}$$

where A is the number of bits set in bit string A; B is the number of bits set in bit string B; and C is the number of bits set common to both bit strings. The coefficients are not generally interconvertable; however, in the extreme case, when bit string B has no bits set that are not set in A, we can write

$$A = C + E$$

where E is the number of bits set uniquely in A, and $B = C$.

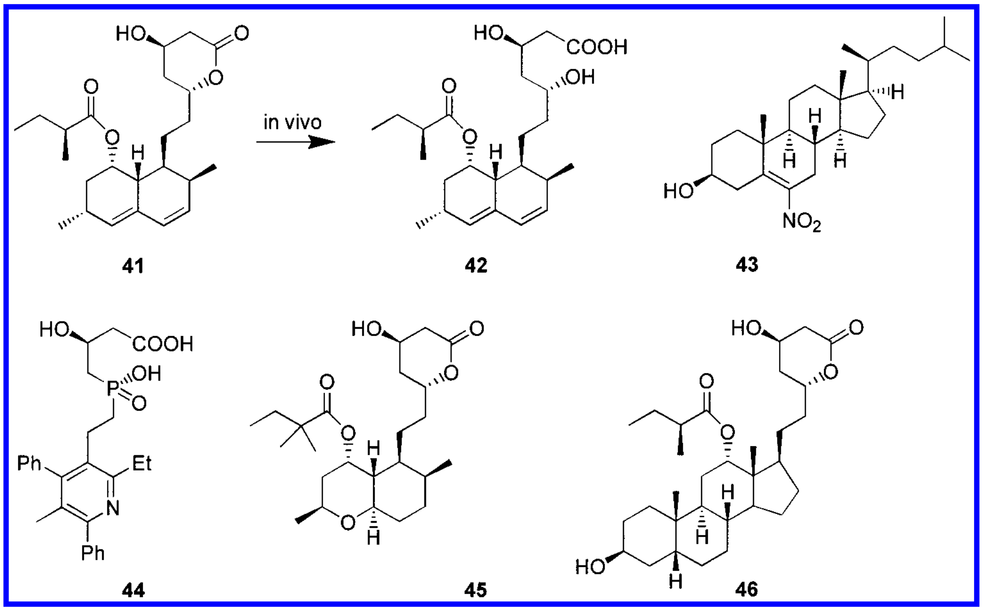


Figure 12. HMG-CoA reductase inhibitors.

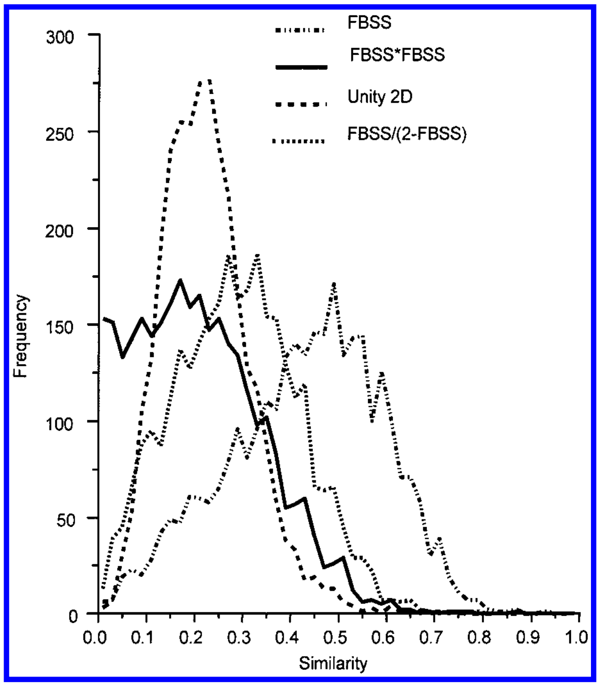


Figure 13. Scaling the rigid FBSS all fields similarities of the random pairs.

In this case, the coefficients simplify to

$$S_{\text{cosine}} = \frac{C}{\sqrt{C^2 + CE}} = \sqrt{\frac{C}{C+E}}$$

$$S_{\text{Tanimoto}} = \frac{C}{C+E} = S_{\text{cosine}}{}^2$$

On the other hand, if we assume that each bit string differs from the maximum common substring by the same number of bits, then

$$A = B = C + F$$

where F is the number of bits in A and B that differ from C, and now the coefficients simplify to

$$S_{\text{cosine}} = \frac{C}{\sqrt{(C+F)(C+F)}} = \frac{C}{C+F} \Rightarrow \frac{C}{F} = \frac{S_{\text{cosine}}}{1 - S_{\text{cosine}}}$$

$$S_{\text{Tanimoto}} = \frac{C}{C+2F} \Rightarrow \frac{C}{F} = \frac{2S_{\text{Tanimoto}}}{1 - S_{\text{Tanimoto}}}$$

By combining the above two equations

$$S_{\text{Tanimoto}} = \frac{S_{\text{cosine}}}{2 - S_{\text{cosine}}}$$

The effects of both transformations on the FBSS All similarities for the random pairs is shown in Figure 13. Of the two extreme cases discussed above, the second is the more probable, and it can be seen that the general shape of the distribution of $S_{\text{FBSS}}/(2 - S_{\text{FBSS}})$ is more similar to the distribution of the UNITY similarities than is the general shape of the $S_{\text{FBSS}}*S_{\text{FBSS}}$ plot; the latter is, however, closer to the distribution of the UNITY similarities in the higher similarity range, which is, of course, the most important part of the distribution for a similarity searching method.

We can combine the two types of similarity measure in various ways. Here, we have calculated the dissimilarity distance D as

$$D = \sqrt{(1 - S_{\text{FBSS}}{}^2)^2 + (1 - S_{\text{UNITY}})^2}$$

(so that $0 \leq D \leq \sqrt{2}$) for the BIOSTER pairs and random pairs reported in Table 1. The average similarities are 0.71 ± 0.24 and 1.11 ± 0.12, respectively, which means that 73% of the BIOSTER pairs have a dissimilarity significantly less (i.e., more than two standard deviations less) than random: this percentage is higher than any of the corresponding values that can be calculated from the data in Table 1, which relates to the individual similarity measures. It would thus seem that combining the two types of similarity measure gives a higher degree of discrimination than using the individual measures. Again, the frequency distributions of D have been partitioned by molecular size. The result is shown in Figure 14, where it will be seen that the combined measure D shows no significant size dependency as a result of the size dependencies of FBSS and UNITY neutralizing each other.

We have tested the effectiveness of the combined dissimilarity, D, relative to the individiual measures by carrying out a set of four searches of the entire BIOSTER database. The ten most similar structures, for S_{FBSS} and S_{UNITY}, and the ten least dissimilar structures, in the case of D, (from here on called “top ten”) were examined to see how many compounds were found to exhibit the same activity as the target. The results are summarized in Table 5. The absolute numbers of hits found for the different targets cannot be compared directly, since there are different numbers of actives in each of the activity classes.

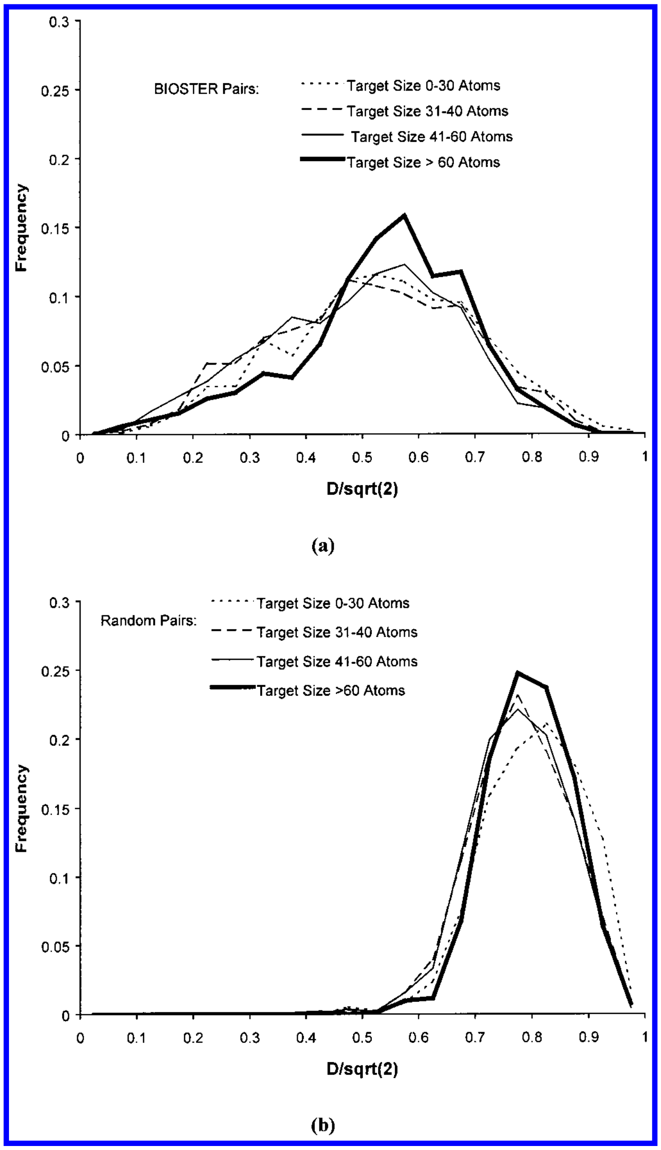
The first search used isoprenaline (**2**, see Figure 6) as the target structure for adrenergic agonists. Only four actives were among the FBSS top ten hits, and six were found among the UNITY and D (**2**, **7**, **1**, **3**, **4** and **8**, in order of decreasing similarity) top ten. The target compound is β-selective, but only a few of the hits have the correct β-selectivity: two in the case of FBSS (**2** and **4**) and UNITY (**2** and **8**), and three when using D (**2**, **4**, and **8**). However, many adrenergic substances in BIOSTER are either not subtype specific or the specificity is not given, so that only one compound (**7**) in the top ten of UNITY and D is specified as an α-agonist.

Ampicillin (**11**; see Figure 7) was chosen as the target structure for alanine-alanine transpeptidase inhibitors, and in all cases the top ten structures also belong to the same activity class regardless of the measure used. Most of the actives retrieved have the typical penicillin skeleton, and where they do have a different scaffold, they are typically cephalosporins (e.g., **13**−**15**). FBSS does retrieve a sulfur-free compound, the carbacephalosporin (**17**). To find a compound from a different lead series, we excluded all structures from the database that have the same core skeleton as compound **11** (since these are always retrieved as most similar). When this was done, UNITY finds only four actives in the top ten structures (structures **20**, **21**, **13**, and **14**, in order of decreasing similarity), FBSS retrieves five actives (structures **13**, **17**, **14**, **16**, and **15**, in order of decreasing similarity); and the combination of UNITY and FBSS results in the retrieval of seven actives. The seven actives (structures **13**, **14**, **16**, **17**, **15**, **18**, and **19**, in order of decreasing

Table 5. Results from Searching the Entire BIOSTER Data Set

target	structures excluded from BIOSTER data set	specification of intrinsic activity or selectivity[a]	N[b]	hits in top ten of		
				FBSS[c]	UNITY	D[d]
isoprenaline **2**	–	β agonists	8	2	2	3
	–	all β binding structures	17	2	2	3
ampicillin **11**	–	–	75	10	10	10
	structures with penicillin skeleton	–	57	5	4	7
estradiol **30**	–	agonist only	18	6	6	8
	–	–	32	7	8	9
	structures with estrogen skeleton	agonist only	13	4	5	6
	structures with estrogen skeleton	–	22	5	5	7
	–	–	46	5	6	6
lovastatine **42** (open form)	–	–	11	2	4	4

[a] If there is no further specification, all compounds belonging to the same activity class are regarded as hits. [b] Possible number of hits. [c] FBSS using all fields and rigid structures. [d] D is the similarity measure derived by combining FBSS (all fields) and UNITY 2D similarities.



(a)

(b)

Figure 14. Frequency distributions of D for different molecular sizes within (a) BIOSTER pairs and (b) random pairs.

similarity) are mainly cephalosporins, but other sulfur-free structures are also found (**16** and **17**) as well as one structure **18** which is not a β-lactam.

The third search was for analogues of estradiol (**30**; see Figure 10). In most cases the top ten structures consisted of compounds in the correct activity class; however, some of the compounds retrieved are inhibitors (e.g., **35**), rather than analogues. The largest number of agonsits are found using the combined method (D). Many of the hits have the estrogen skeleton. Exceptions are **32**, which is found by all measures, **33**, which is found with FBSS and D, **34**, which is found only with FBSS, and **31**, which is found by UNITY and D. When the search was performed after excluding all compounds that contain the estrogen skeleton, five actives were found as hits with UNITY (structures **32**, **31**, **33**, **34**, and **36**, in order of decreasing similarity) and FBSS (structures **35**, **33**, **32**, **37**, and **31**, in order of decreasing similarity) alone, the combination D is again more effective and retrieves seven actives (structures **32**, **33**, **31**, **34**, **35**, **36**, and **37**, in order of decreasing similarity).

Searching for **42** (see Figure 12) as an HMG-CoA reductase inhibitor resulted in only two hits retrieved by FBSS (**45**, **42**) with the target itself not being the most similar structure. This reflects the nondeterministic nature of the GA that underlies FBSS, which does not guarantee that the best alignment will be always found. UNITY and D both retrieved the structures **42**, **45**, **41**, and **46**.

In the second series of experiments, a comparable set of four searches was performed in which fusion was effected by combining the ranks (rather than the similarity scores themselves).[41] Fusion based on ranks can be performed directly without the need for scaling as is required when fusing the similarity scores themselves. The results for rank-based fusion were found to be very similar to those already reported; however, there are differences in terms of efficiency which can be significant. Although combining ranks is simpler to calculate than the distance measure, D, that was used for fusing the similarity scores, D has the advantage of being more computationally efficient since it can be possible to reduce the total number of calculations required quite considerably. From the definition of D it can be seen that D can only be smaller than a given threshold t if the following minimum requirements are fulfilled:

$$(1 - S_{\mathrm{UNITY}}) < t \quad \text{and} \quad (1 - S_{\mathrm{FBSS}}^{2}) < t$$

To find all structures within a given threshold t of a target structure, it is necessary to calculate the UNITY similarity for all compounds, but then only those compounds for which $(1 - S_{\mathrm{UNITY}}) < t$ is true need to be compared using the FBSS measure. Since the calculation of S_{UNITY} is much faster than the calculation of S_{FBSS}, the calculation time should be reduced significantly.

CONCLUSIONS

In this paper we have considered the effectiveness of various similarity measures for identifying pairs of molecules that are bioisosteres. Experiments with measures based on UNITY 2D fingerprints and on molecular field information using FBSS demonstrate clearly the very different ways in which these two types of measure are able to identify intermolecular structural similarities. For example, the UNITY measure is very sensitive to heteroatom substitutions whereas FBSS is concerned with the fields projected around atoms rather than the element types themselves; conversely the UNITY measure is independent of 3D conformation, whereas FBSS is very dependent on the accuracy of the 3D conformations used which can be a problem for large flexible structures. The very different characteristics of the two methods suggest that superior performance might be obtained by combining them in some way, and this is demonstrated in our data fusion experiment. We conclude that each type of similarity measure is able to identify at least some type of bioisosteric relationship, and that this ability is maximized by taking account of more than a single type of similarity relationship.

ACKNOWLEDGMENT

We thank the DAAD (German academic exchange service) for the award of a NATO scholarship to A.S., Synopsys Scientific Systems for provision of the BIOSTER database, and Tripos Inc. for software support.

REFERENCES AND NOTES

(1) *Concepts and Applications of Molecular Similarity*; Johnson, M. A., Maggiora, G. M., Eds.; John Wiley: New York; 1990.

(2) Patterson, D. E.; Cramer, R. D.; Ferguson, A. M.; Clark, R. D.; Weinberger, L. E. Neighbourhood Behaviour: A Useful Concept for Validation of "Molecular Diversity" Descriptors. *J. Med. Chem*. **1996**, *39*, 3049−3059.

(3) *Molecular Similarity in Drug Design*; Dean, P. M., Ed.; Chapman and Hall: Glasgow, 1994.

(4) Willett, P.; Barnard, J. M.; Downs, G. M. Chemical Similarity Searching. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 983−996.

(5) Downs, G. M.; Willett, P. Similarity Searching in Databases of Chemical Structures. *Rev. Comput. Chem.* **1995**, *7*, 1−66.

(6) Dean, P. M. In *Designing Bioactive Molecules: Three-Dimensional Techniques and Applications*; Martin. Y. C., Willett, P., Eds.; American Chemical Society: Washington, DC, 1998.

(7) Pepperrell, C. A.; Willett, P.; Taylor, R. Implementation and Use of an Atom-Mapping Procedure for Similarity Searching in Databases of 3-D Chemical Structures. *Tetrahedron Comput. Methodol.* **1990**, *3*, 575−593.

(8) Bemis, G. W.; Kuntz, I. D. A Fast and Efficient Method for 2D and 3D Molecular Shape Description. *J. Comput.-Aided Mol. Des.* **1992**, *6*, 607−628.

(9) Fisanick, W.; Cross, K. P.; Rusinko, A. Similarity Searching on CAS Registry Substances. 1. Global Molecular Property and Generic Atom Triangle Geometric Searching. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 664−674.

(10) Nilakantan, R.; Bauman, N.; Venkataraghavan, R. A New Method for Rapid Characterisation of Molecular Shape: Applications in Drug Design. *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 79−85.

(11) *3D QSAR in Drug Design;* Kubinyi, H., Folkers, G., Martin, Y. C., Eds.; Kluwer/ESCOM: Leiden, 1998.

(12) Carbó, R.; Leyda, L.; Arnau, M. How Similar is a Molecule to Another? An Electron Density Measure of Similarity Between Two Molecular Structures. *Int. J. Quantum Chem.* **1980**, *17*, 1185−1189.

(13) Richard, A. M. Quantitative Comparison of Molecular Electrostatic Potentials for Structure−Activity Studies. *J. Comput. Chem.* **1991**, *12*, 959−969.

(14) Good, A. C.; Hodgkin, E. E.; Richards, W. G. The Utilization of Gaussian Functions for the Rapid Evaluation of Molecular Similarity. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 188−191.

(15) Sanz, F.; Manaut, F.; Rodriguez, J.; Lozoya, E.; Lopez-de-Brinas, E. MEPSIM: a Computational Package for Analysis and Comparison of Molecular Electrostatic Potentials. *J. Comput.-Aided Mol. Des.* **1993**, *7*, 337−347.

(16) Petke, J. D. Cumulative and Discrete Similarity Analysis of Electrostatic Potentials and Fields *J. Comput. Chem.* **1993**, *14*, 928−933.

(17) Mestres, J.; Rohrer, D. C.; Maggiora, G. M. A Molecular-Field-Based Similarity Study of Non-Nucleoside HIV-1 Reverse Transcriptase Inhibitors. *J. Comput.-Aided Mol. Des.* **1999**, *13*, 79−93.

(18) Wild, D. J.; Willett, P. Similarity Searching in Files of Three-Dimensional Chemical Structures: Alignment of Molecular Electrostatic Potential Fields with a Genetic Algorithm. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 159−167.

(19) Thorner, D. A.; Wild, D. J.; Willett, P.; Wright, P. M. Similarity Searching in Files of Three-Dimensional Chemical Structures: Flexible Field-Based Searching of Molecular Electrostatic Potentials. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 900−908.

(20) Drayton, S. K.; Edwards, K.; Jewell, N. E.; Turner, D. B.; Wild, D. J.; Willett, P.; Wright, P. M.; Simmons, K. Similarity Searching in Files of Three-Dimensional Chemical Structures: Identification of Bioactive Molecules. *Internet J. Chem.* at URL http://www.ijc.com/articles/1998v1/37/.

(21) Goldberg, D. E. *Genetic Algorithms in Search, Optimization and Machine Learning*; Addison-Wesley: Wokingham, 1989.

(22) *Handbook of Genetic Algorithms*; Davis, L., Ed.; Van Nostrand Reinhold: New York; 1991.

(23) *Handbook of Evolutionary Computing*; Back, T., Fogel, D., Michalewicz, Z., Eds.; Oxford University Press USA: New York, 1997.

(24) Clark, D. E. *Evolutionary Algorithms in Computer-Aided Molecular Design*; at URL http://panizzi.sheffield.ac.uk/cisrg/links/ea_bib.html.

(25) The World Drugs Index is available from Derwent Information at URL http://www.derwent.co.uk/

(26) The CONCORD, SYBYL, and UNITY programs are available from Tripos Inc. at URL http://www.tripos.com.

(27) Lipinski, C. A. Bioisosterism in Drug Design. *Annu. Rep. Med. Chem.* **1986**, *21*, 283−291.

(28) Burger, A. Isosterism and Bioisosterism in Drug Design. *Prog. Drug Res*. **1991**, *37*, 288−371.

(29) Ujváry, I. BIOSTER−A Database of Structurally Analogous Compounds. *Pestic. Sci.* **1997**, *51*, 92−95.

(30) The BIOSTER database is available from Synopsys Scientific Systems at URL http://www.synopsys.co.uk/

(31) The MDL data formats are available at http://www.mdli.com/

(32) Stewart, J. J. P. MOPAC: a Semiempirical Molecular Orbital Program. *J. Comput.-Aided Mol. Des.* **1990**, *4*, 1−105.

(33) The CORINA program is available from Oxford Molecular Group at URL http://www.oxmol.co.uk/

(34) Allen, F. H.; Davies, J. E.; Galloy, J. J.; Johnson, O.; Kennard, O.; Macrae, C. F.; Mitchell, E. M.; Mitchell, G. F.; Smith, J. M.; Watson, D. G. The Development of Versions 3 and 4 of the Cambridge Structural Database System. *J. Chem. Inf. Comput. Sci.* **1991**, *31*, 187−204.

(35) Flower, D. R. On the Properties of Bit String-Based Measures of Chemical Similarity. *J. Chem. Inf. Comput. Sci*. **1998**, *38*, 379−386.

(36) Johnson, J. W.; Ascher, P. Glycine Potentiates the NMDA Response in Cultured Mouse-Brain Neurons. *Nature* **1987**, *325*, 529−531.

(37) Mallamo, J. P.; Earley, W. G.; Kumar, V.; Subramanyam, C.; Dority J. A.; Miller, M. S.; Dehavenhudkins, D. L.; Ault, B.; Herrmann, J. L.; Dung, J. S.; McMullen, L. A.; Jaeger, E.; Kullnig, R.; Magee, L. J. Identification, Synthesis and Characterization of a Unique Class of *N*-Methyl-D-Aspartate Antagonists−The 6,11-Ethanobenzo[*B*]Quinolizinium Cation. *J. Med. Chem.* **1994**, *37*, 4438−4448.

(38) Hall, D. L. *Mathematical Techniques in Multisensor Data Fusion*; Artech House: Northwood, MA, 1992.

(39) *Proceedings of the International Conference on Multisource-Multisensor Information Fusion*, *Fusion '98*, Arabnia, H. R., Zhu, D., Eds.; CSREA Press: 1998.

(40) Ginn, C. M. R.; Turner, D. B.; Willett, P.; Ferguson, A. M.; Heritage, T. W. Similarity Searching in Files of Three-Dimensional Chemical Structures: Evaluation of the EVA Descriptor and Combination of Rankings Using Data Fusion. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 23−37.

(41) Ginn, C. M. R.; Willett, P.; Bradshaw, J. Combination of Molecular Similarity Measures Using Data Fusion. *Perspect. Drug Discovery Des.*, in press.

CI990263G