Combination Of Molecular Similarity Measures Using Data Fusion

Claire M. R. Ginn and Peter Willett*

University of Sheffield, Sheffield S10 2TN, UK

John Bradshaw

GlaxoWellcome Research and Development Limited, Stevenage, SG1 2NY, UK

Abstract Many different measures of structural similarity have been suggested for matching chemical structures, each such measure focusing upon some particular type of molecular characteristic. The multi-faceted nature of biological activity suggests that an appropriate similarity measure should encompass many different types of characteristic, and this paper discusses the use of data fusion methods to combine the results of searches based on multiple similarity measures. Experiments with several different types of dataset and activity suggest that data fusion provides a simple, but effective, approach to the combination of individual similarity measures. The best results were generally obtained with a fusion rule that sums the rank positions achieved by each molecule in searches using individual measures.

* To whom all correspondence should be addressed. Email: p.willett@sheffield.ac.uk

INTRODUCTION

Measures of inter-molecular similarity play an important role in drug- and pesticide-discovery programmes, being used for both database searching [1] and structure-activity studies [2]. Many different types of similarity measure have been described in the literature (see, *e.g.*, [3-5]) but the great majority of published studies have considered the use of only a single type of similarity measure: in many cases, indeed, a description of a new type of similarity measure forms the principal focus of the publication. Even where this is not the case, multiple measures have typically been employed only as the input to a comparative study that seeks to identify the "best" measure, using some quantitative performance criterion. As an example, an early study in our laboratory [6] compared 36 different similarity measures by means of simulated leave-one-out property prediction, and concluded that the Tanimoto coefficient was the most appropriate

similarity coefficient of those tested for measuring the resemblances between pairs of fragment bit-strings. Such comparisons, of which there are many in the literature, are limited in that they assume, normally implicitly, that there is some specific type of structural feature, weighting scheme or whatever that is uniquely well suited to describing the type(s) of biological activity that are being sought for in a similarity search. The assumption cannot be expected to be generally valid, given the multi-faceted nature of biological activities, and this paper investigates the use of *data fusion* [7] for combining multiple similarity measures.

DATA FUSION

Background Data fusion is "a process of combining inputs from sensors with information from other sensors, information processing blocks, databases, or knowledge bases, into one representational format" [8]. Defence applications have provided much of the driving force for the development of data fusion techniques, with published examples including establishing the friend-or-foe nature of an incoming missile or aeroplane, predicting the range and direction of a battlefield target, and navigating an un-manned armoured vehicle. Other applications include surveillance operations by law enforcement agencies, real-time control of continuous manufacturing processes, the provision of all-weather visibility for aircraft pilots, and multi-imaging systems for the analysis of medical images (see, *e.g.*, [9]). However, data fusion can be, and is, used in much more commonplace situations: for example, establishing that it is safe to cross a road involves taking input from one's ocular sensors (eyes) and aural sensors (ears), and then combining this information with the knowledge that an empty road is a safe road to give an output denoting the safety of the proposed action. Again, a committee in which all members can contribute will often arrive at a superior decision to the one that would have been reached by just the committee chair - although there are, of course, many exceptions to such a rule!

The basic rationale for data fusion is that using the information presented by a number of sensors enables further information to be inferred that would be outside the capabilities of a single sensor. For example, if one sensor detects a tank, then all that can be deduced is the existence and the position of that tank. However, if two sensors detect the same tank then inferences can be made regarding the direction of its movement, while the addition of a temporal dimension permits the tank's velocity to be calculated. Add to that the ability to compare the observed behaviour with records of past behaviour of tanks and the system becomes capable of threat analysis. As well as being able to infer more information, the use of a fusion system also leads to both qualitative and

quantitative improvements in several ways. Thus, improved operational performance can occur if one of the sensors were to become damaged, as there would still be information coming in from the others (an obvious advantage in military applications where sensors will be exposed to combat conditions and are thus liable to become damaged). Data fusion leads to extended coverage since multiple sensors can cover disparate areas, times and qualities, and it leads to an increased level of confidence in the results since multiple sensors can act together to confirm an event and to reduce any ambiguity surrounding, *e.g.*, the classification of an event.

Combination of rankings Our interest in data fusion methods arose from recent work on their application to information retrieval (IR), specifically to the combination of the rankings produced by different retrieval mechanisms when applied to databases of textual documents. An early study is that by Belkin *et al.* [10], in which data fusion was used to combine the results of a series of searches of bibliographic databases, conducted in response to a single query, but employing different indexing and searching strategies. A query was processed using different strategies, each of which was used to produce a ranking of a set of documents in order of decreasing similarity with the query. The ranks for each of the documents were then combined using one of several different fusion rules (including the MIN, MAX and SUM rules discussed below); the output of the fusion rule was taken as the document's new similarity score and the fused lists were then re-ranked in descending order of similarity. This work soon led to many other studies (see, *e.g.*, [11-13]) and the combination of document rankings is now a well-established technique, as is exemplified by its use in a meta-search engine that provides access to the World Wide Web using a combination of different search engines [14].

The work on chemical data fusion reported here is based directly on these previous IR studies, and involves the simple procedure shown in Schema 1, where a user-defined target structure is searched against a database using several different similarity measures. The fusion rules that we use here are based on those identified by Belkin *et al.* [10], and are summarised in Table 1. It will be seen that the MIN and MAX rules represent the assignment of extreme ranks to database structures and it is thus hardly surprising that both can be highly sensitive to the presence of a single "poor" retrieval system amongst those that are being combined. The SUM rule is expected to be more stable against the presence of a single poor or noisy input ranking; here, each database structure is assigned the sum of all the rank positions at which it occurs in the input lists. This report considers just these three rules but there are clearly many others that could be considered, *e.g.*, the median, the product, the harmonic mean, *etc.* of the individual rankings.

The combined scores output by the fusion rule are then used to re-order the database structures to give the final ranked output. In many cases, especially with the SUM rule, the application of the fusion rule may result in the assignment of the same score to two or more items. When this happens, it is necessary to specify a further sort key to allow the resolution of the tied structures, *e.g.*, alphabetical ordering of the canonicalised connection tables describing the tied database structures or the allocation of weights to individual rankings (perhaps based on past performance in similarity searches) so that a high position in one ranking would differ in importance from that same position in another ranking.

Chemical applications Chemical applications of data fusion are not completely novel. As long ago as 1973, Clerc and Erni noted that "when data from several different spectroscopic methods are used for comparison purposes, greatly enhanced performance may be expected because the methods complement each other" [15] and went on to discuss the use of a scoring scheme based on weighted contributions from each of several molecular properties and spectra. More recently, Masui and Yoshida [16] have reported the use of the SPECTRA system for combining the similarity scores obtained in searches of a database containing mass, IR, and ^{1}H and ^{13}C NMR spectral data when one or more of the spectra are missing for a particular sample molecule. In work more analogous to that reported here, Kearsley *et al.* have used both similarity-based and rank-based procedures to combine pairs of similarity searches of the Standard Drug File database, and found that significant improvements in performance could be achieved in simulated property prediction experiments [17, 18]. Finally, So and Karplus have recently advocated combining different QSAR methods to obtain models with heightened predictivity [19].

Our initial studies of data fusion were undertaken as part of a project to evaluate the EVA descriptor, which characterises a molecule by its fundamental vibrational fingerprint [20]. Although originally developed for QSAR applications, the EVA descriptor can also be used for similarity searching and a range of EVA-based similarity measures were hence evaluated using a dataset containing 8178 molecules from the Starlist file [21]. Comparable searches were also carried out using the 2D similarity searching routines in the UNITY chemical information management system [22], and using data fusion to combine the two individual types of ranking. Simulated leave-one-out property prediction experiments using the logP data in the Starlist file showed that, on average, the fused rankings appeared to be better than the original 2D and EVA rankings. Although the differences were not always statistically significant, the study provided at

least some evidence that data fusion could be used to improve the performance of similarity searching in chemical databases: the remainder of this paper reports further experiments that have been undertaken to ascertain the accuracy of this conclusion. Full details of the work are provided by Ginn [23].

CELLULAR-UPTAKE DATASET

The dataset These experiments involved a set of 136 biological dyes that are used to stain cells so as to visualise various organelles, specifically the lysosomes (L), the mitochondria (M), and the nucleus (N). These three broad activity classes were subdivided into eight mechanism-specific subclasses [24] and each molecule in the dataset was allocated an 8-bit activity bit-string in which the *i*-th bit was switched on if the molecule exhibited the *i*-th activity. Three different types of descriptor were used to characterise the molecules in this dataset: 2D fragments, 3D fragments and physical properties. The 2D fragment descriptors used here were the fingerprints produced by Barnard Chemical Information Limited (BCI) [25], while the 3D fingerprints were based on the NBN non-bonded torsion angle descriptor developed by Bath *et al*. [26]. The physical property descriptor comprised three standardised properties for each molecule: the logarithm of the octanol/water partition coefficient, the net electric charge, and the number of bonds included in the delocalised electron system of the molecule [24]. Each molecule in the dataset was considered as a target for similarity searching using each of the three similarity measures, with the similarity between a pair of molecules being calculated using the Tanimoto coefficient (the simple binary form of this coefficient for the 2D and 3D fingerprint measures and the generalised, non-binary form for the physical property measure) [5]. The three rankings for each target structure were fused using the SUM, MIN and MAX fusion algorithms defined previously.

Comparison of ranks and similarities An inspection of Scema 1 shows that Step 2 of the basic fusion procedure involves the rank positions for each database structure, rather than the similarity scores that are output by the similarity measure. On first sight, the former might seem to be the less intuitively reasonable approach as it involves a loss of information when compared with the use of scores. However, there are two factors associated with the use of similarity scores that lessen their attractiveness. Firstly, as researchers are more likely to be concerned with some number of nearest-neighbours to the target structure, rather than with those items that are above some threshold of similarity, it seems logical to consider the rank positions of the items irrespective of their similarity scores. Secondly, and more importantly, despite having the same

range of scores (such as zero to unity for the binary version of the Tanimoto coefficient [5]), the distributions within these ranges given by different similarity measures may not be directly comparable, with the possibility of biasing the fusion rule in much the same way as unstandardised numeric data can affect the results of a multivariate analysis.

We have compared the distributions of scores for each similarity method at each rank *n*, using the Kolmogorov-Smirnov test, which provides a simple and direct way of testing whether two distributions differ in any way, *e.g.*, in location, dispersion or skewness [27]. If the distributions of the similarity scores for two original similarity measures are significantly different then it would be unwise to fuse them without applying some form of standardisation procedure (*i.e.*, the use of rank positions in the present context). Figure 1 shows plots of the mean similarity scores (averaged over all 136 target structures) at each rank position, n ($1 \leq n \leq 100$). The figure shows that while the 2D and 3D scores are distributed similarly, the physical property scores exhibit a markedly different distribution. Focusing upon the important top parts of the rankings, pairs of the distributions were compared for n=1-10 using the Kolmogorov-Smirnov test: these tests showed that the distribution of scores for the physical properties measure was significantly different ($p \leq 0.01$) to those from both 2D and 3D for n=1-10 and that the distribution of scores for 2D was significantly different to those for 3D for n=1-4. We hence conclude that the distributions of similarity scores can be very different, even if they have the same ranges, thus supporting our use of ranks as the input to the various fusion rules studied here. Similar results were obtained [23] in a comparable study of the EVA and 2D rankings of the Starlist dataset mentioned previously.

Fusion results Having established the appropriateness of rank-based fusion, the main experiments were evaluated in two ways. In the first, a count was made of the molecules ranked in the top ten positions that belonged to the same activity subclass as the target structure. These counts were then averaged over each of the eight subclasses, with the results shown in Table 2, where $L_{1\text{-}4}$ (lysosomes), $M_{1\text{-}2}$ (mitochondria) and $N_{1\text{-}2}$ (nucleus) denote the eight activity subclasses identified in the dataset. The shaded elements denote fusion results that perform at least as well as the best individual similarity measure. It will be seen that the best similarity measure, in terms of actives being highly ranked, varies across activity subclasses; however, the results demonstrate that both SUM and MAX are, overall, to be preferred to the individiual results. SUM also does well if one ranks the measures for each search, rather than using the actual numbers of actives retrieved (which vary considerably from one search to another). For example, in the first row of

Table 2, SUM identifies most actives and is given the rank 1, Phys identifies the next highest number of actives and is given the rank 2 and so on down to 3D, which identifies the smallest number of actives and is thus given the rank 6. The mean ranks obtained in this way, when averaged across the eight activity sub-classes, are listed in the bottom row of the table and demonstrate clearly the effectiveness of the SUM fusion rule with this dataset.

The second set of analyses employed the Hamming distance [5] between the activity bit-strings of the target structure and a database structure, *i.e.*, the number of times that the two bit-strings differ. For example, if the target is active for subclasses L_1, M_1 and N_1 then a Hamming distance of 0 between a database structure and the target indicates that the former is also active in subclasses L_1, M_1 and N_1 and only in those classes, and would thus be a most appropriate hit for that target molecule. Figure 2 shows the mean Hamming distance for each similarity measure across all 136 target structures at rank n ($1 \leq n \leq 10$), and it can be seen that the SUM and MAX fusion algorithms give results that are consistently better (*i.e.*, a smaller mean Hamming Distance) than those from any of the individual similarity methods. A pairwise comparison of similarity methods was carried out using the Wilcoxon Matched-Pairs Signed-Ranks test [27]. Specifically, the test was used to compare the Hamming distances for each fusion rule with each of the original similarity methods, target by target, and thus to indicate whether the two methods that are being compared are significantly different. Table 3 shows the p values for n = 1-10. It can be seen that SUM is significantly better than each of three original similarity methods for all values of n, with 28 out of the 30 sets of comparisons being highly significant ($p \leq 0.01$). MAX also performs well, but MIN is noticeably inferior to the other two fusion rules for this dataset.

Taken together, these results show that the fused similarity measures can, in some cases at least, enable better predictions to be made of the cell-staining activities of the molecules than can the original measures, with SUM appearing to perform best of the three fusion rules tested here. When we take account of the rather variable performance of the individual similarity measures from one activity to another, it can be concluded that SUM-based fusion provides an effective way of generating a reliable single ranking with respect to both a single activity and the activity classes as a whole.

WORLD DRUG INDEX DATASET

Having demonstrated the potential of data fusion on a small dataset, the next set of experiments used a file of structures and associated broad-class bioactivity data from the World Drug Index (WDI) database [28]. Three different types of similarity measure were used here, these being based on 2D fragment occurrence data, 3D geometric information and on molecular fields. The 2D rankings were obtained using the UNITY fingerprints mentioned previously, while the 3D rankings were obtained using the atom-mapping measure described by Pepperrell *et al*. [29]. This measure uses inter-atomic distance information to identify pairs of atoms, one in the target structure and one in the current database structure, that are surrounded by similar patterns of atoms; these initial atomic equivalences are then used to construct an approximate mapping of the target structure onto the database structure. The field-based rankings were obtained using the FBSS (for field-based similarity searching) program described by Drayton *et al.* [30], in which a target structure is aligned with a database structure by means of their steric, hydrophobic and electrostatic fields. The particular version of the program used here considered all three types of field in the generation of an alignment, and hence in the resulting similarity score (this corresponding to the 'All' search of Drayton *et al*. [30]).

Ten target structures were chosen that had been used previously by Kearsley *et al*. in their studies of WDI-based similarity searching [17]. The similarity searches were performed on datasets of approximately 3600 structures, each containing the activity class for the target structure with an additional 3500 randomly-selected WDI molecules. The data available for fusing comprised of three sets of rankings (one for each of the original similarity measures) for each of the ten targets, with the effectiveness of each search being measured by the number of molecules in the top-50 rank positions that had the same activity as the target; other performance measures for this dataset are discussed by Ginn [23]. Table 4 lists the numbers of actives identified in the original and fused searches for each of the 10 target structures. The results obtained are similar to those obtained with the cellular-uptake dataset: while the fused results are not always as good as the best individual result, they provide a generally high, and thus robust, level of effectiveness whereas the best original measure varies from target to target. This is particularly clear if one inspects the mean activities and ranks at the bottom of the table, where it will be seen that SUM would again seem to be the fusion rule of choice.

KAHN DATASET

The dataset The final section evaluates data fusion when a larger number of original similarity measures is available. The dataset used here is described by Kahn in a discussion of descriptors for the analysis of combinatorial libraries [31]: it contains 75 compounds each belonging to one of 14 well-defined activity classes (angiotensin-converting enzyme inhibitors, acetylcholine receptor inhibitors, antagonists of 2-aminoproprionic acid, aldose reductase inhibitors, angiotensin-II receptor antagonists, beta adrenergic blockers of the type-3 receptor, cyclo oxygenase 2 receptor antagonists, dopamine 3 receptor (ant)agonists, endothelin receptor (ant)agonists, histamine 2 antagonists, neurokinase-1 receptor antagonists, HIV-1 protease inhibitors, non-nucleoside HIV reverse transcriptase inhibitors, and steroid aromatase inhibitors).

Six similarity measures were used to generate rankings: the Molecular Simulations Inc. (MSI) [32] Jurs descriptors; FBSS (as discussed in the previous section); two types of ChemX 3D flexible fingerprints [33]; and two types of Daylight 2D fingerprints [34]. The Jurs descriptors are part of the MSI Cerius^2 package, and describe shape and electronic charge by mapping the atomic partial charges onto the solvent accessible areas of the individual atoms within a molecule. All of the 30 possible Jurs descriptors [35] were calculated for each member of the dataset. The values were then normalised, and the similarity between pairs of sets of values calculated using the non-binary Tanimoto coefficient. In what follows, the inclusion of the Jurs rankings in a fusion combination is indicated by “J”. The FBSS similarity measure has been described previously: its inclusion in a fusion combination is denoted by “F”. The ChemX 3D flexible fingerprint keys record the presence or absence of potential pharmacophoric patterns (consisting of three pharmacophore centres and the associated inter-atomic distances) in any of the low-energy conformations identified by a rule-based conformational analysis of a molecule. Two sets of similarity scores were generated from these fingerprints: the Tanimoto coefficient scores and the Tversky similarity scores [5, 36], the inclusion of these in a fusion combination being denoted by “3” or by “T”, respectively. The Daylight fingerprints were based on unfolded fingerprints considering pathlengths of up to 7, the inclusion of these in a fusion combination being denoted by “2” (for a standard fingerprint where a bit is either set or not set) or by “N” (for a fingerprint where a count is kept of how many times each bit is set), respectively. Thus 23F, for example, represents the fusion of the standard Daylight, Tanimoto ChemX and FBSS rankings. The similarity scores for these experiments were calculated using either the binary or non-binary versions of the Tanimoto coefficient, as appropriate.

Fusion results In view of its performance in the studies discussed above, we used just the SUM rule for the fusion experiments, with all possible combinations of rankings from the similarity methods being studied (in much the same way as So and Karplus have very recently evaluated the effectiveness of all possible combinations of seven different QSAR methods [19]). Table 5 details the mean numbers of actives (*i.e.*, molecules with the same activity as the target structure) found in the top-10 nearest neighbours when averaged over all 75 target structures. The values of *c* at the top of the table denote the number of similarity measures that were fused (so that, *e.g.*, *c*=1 represents the original measures and *c*=2 represents the fusion of a pair of the original measures) and a shaded element indicates a fused combination that is better than the best original individual measures (which was the ChemX keys with the Tanimoto coefficient).

It will be seen that very many of the fused combinations in Table 5 are shaded, thus providing further support for the use of SUM to fuse similarity rankings, and Ginn reports similar results from other analyses of this dataset [23]. The table also shows that the fraction of the combinations that are shaded increases in line with *c*, so that all combinations with $c \geq 4$ perform at least as well as the best of the individual similarity measures. However, it is not the case that, *e.g.*, the *c*=5 combinations are invariably superior to the *c*=4 combinations, and the best result overall was obtained with 23FJT (rather than with 23FJNT, the combination involving all of the individual measures). Thus, while simply fusing as many individual measures as are available in a similarity investigation would appear to perform well, superior results may be obtained from fusing a subset of the individual measures; this has also been noted in searches of text databases [10] but there is no obvious predictive mechanism for identifying an optimal combination *a priori* [23, 37].

CONCLUSIONS

In this paper we have discussed the use of data fusion methods to combine the rankings resulting from similarity searches of chemical datasets. Our experiments, which have employed a range of types of molecule and performance criterion, demonstrate that use of a fusion rule such as SUM will generally result in a level of performance (however this is quantified) that is at least as good (when averaged over a number of searches) as the best individual measure: since the latter often varies from one target structure to another in an unpredictable manner, the use of a fusion rule will generally provide a more consistent level of searching performance than if just a single similarity measure is available.

Acknowledgements We thank the following: the Engineering and Physical Sciences Research Council and GlaxoWellcome Research and Development Limited for funding; BioByte Corp., Darren Green, Richard Horobin, Sonia Ranade, David Turner and David Wild for data; and Barnard Chemical Information Limited and Tripos Inc. for software support.

REFERENCES

1. Downs, G.M. and Willett, P., Rev. Comput. Chem., 7 (1995) 1.
2. Dean, P.M. and Perkins, T.D.J., In: Martin. Y.C. and Willett, P. (Eds.). Designing Bioactive Molecules: Three-Dimensional Techniques and Applications, American Chemical Society, Washington DC, 1998, pp. 199-218.
3. Special issue devoted to molecular similarity, J. Chem. Inf. Comput. Sci., 32 (1992) 577-752.
4. Dean, P.M. (editor), Molecular Similarity in Drug Design, Chapman and Hall, Glasgow, 1975.
5. Willett, P., Barnard, J.M. and Downs, G.M., J. Chem. Inf. Comput. Sci., 38 (1998) 983.
6. Willett, P. and Winterman, V., Quant. Struct.-Activ. Relat., 5 (1986) 18.
7. Hall, D.L., Mathematical Techniques in Multisensor Data Fusion, Artech House, Northwood MA, 1992.
8. Kokar, M and Kim, K., Control Eng. Pract., 2 (1994) 803.
9. Arabnia, H.R. and Zhu, D. (editors), Proceedings of the International Conference on Multisource-Multisensor Information Fusion, Fusion'98, CSREA Press, 1998.
10. Belkin, N.J., Kantor, P., Fox, E.A. and Shaw, J.B., Inf. Proc. Manag., 31 (1995) 431.
11. Savoy, J., Ndarugendamwo, M. and Vrajitoru, D., Proceedings of the Fourth Text Retrieval Conference, National Institute for Standards and Technology NIST Special Publication 500-236, Gaithersberg MD, 1996, pp 537-547.
12. Lee, J.H., Proceedings of the Twentieth Annual International Conference on Research and Development in Information Retrieval, Association for Computing Machinery, New York, 1997, pp 267-276.
13. Pfeifer, U., Poersch, T. and Fuhr, N., Inf. Proc. Manag., 32 (1996) 667.
14. Smeaton, A.F. and Crimmins, F., URL http://www.inf.udec.cl/~campos/fusion/fusionpc/fusion-www6.html
15. Clerc, T. and Erni, F., Topics Curr. Chem., 39 (1973) 91.
16. Masui, H. and Yoshida, M., J. Chem. Inf. Comput. Sci., 36 (1996) 294.
17. Kearsley, S.K., Sallamack, S., Fluder, E.M., Andose, J.D., Mosely, R.T. and Sheridan, R.P., J. Chem. Inf. Comput. Sci., 36 (1996) 118.
18. Sheridan, R.P., Miller, M.D., Underwood, D.J. and Kearsley, S.K., J. Chem. Inf. Comput. Sci., 36 (1996) 128.
19. So, S.-S. and Karplus, M., J. Comput.-Aid. Mol. Design, 13 (1999) 243.
20. Ginn, C.M.R., Turner, D.B., Willett, P., Ferguson, A.M. and Heritage, T.W., J. Chem. Inf. Comput. Sci., 37 (1997) 23.
21. The Starlist file is available from BioByte Corp. at URL http://clogp.pomona.edu/
22. UNITY is available from Tripos Inc. at http://www.tripos.com
23. Ginn, C.M.R. The Application of Data Fusion to Similarity Searching of Chemical Databases. PhD thesis, University of Sheffield, 1998.
24. Ranade, S.S. Prediction of Cellular Uptake of Foreign Chemicals Using Cluster Analysis, PhD thesis, University of Sheffield, 1998.
25. Barnard Chemical Information Limited is at URL http://www.bci1.demon.co.uk

26. Bath, P.A., Poirrette, A.R., Willett, P. and Allen, F.H., J. Chem. Inf. Comput. Sci., 34 (1994) 141.
27. Siegel, S. and Castellan, N.J., Nonparametric Statistics. McGraw-Hill, New York, 1988.
28. The World Drug Index database is available from Derwent Information at URL http://www.derwent.co.uk
29. Pepperrell, C.A., Taylor, R. and Willett, P., Tetrahedron Comput. Methodol., 3 (1990) 575.
30. Drayton, S.K., Edwards, K., Jewell, N.E., Turner, D.B., Wild, D.J., Willett, P., Wright, P.M. and Simmons, K., Internet J. Chem., URL http://www.ijc.com/articles/1998v1/37/.
31. Kahn, S.D., Combinatorial Libraries: Structure Activity Analysis. In: Encyclopedia of Computational Chemistry, Schleyer, P. v. R.; Allinger, N. L.; Clark, T.; Gasteiger, J.; Kollman, P. A.; Schaefer III, H. F.; Schreiner, P. R. (Eds.); John Wiley: Chichester, 1998, vol 1, 417-425.
32. Molecular Simulations Inc. is at URL http://www.msi.com
33. ChemX products are available from Oxford Molecular Limited at URL http://www.oxmol.co.uk
34. Daylight Chemical Information Systems Inc. is at URL http://www.daylight.com
35. Stanton, D.T. and Jurs, P.C., Anal. Chem., 62 (1990) 2323.
36. Bradshaw, J., URL http://www.daylight.com/meetings/mug97/Bradshaw/MUG97/tv_tversky.html
37. Smeaton, A.F., Proceedings of the Twenty-Second Annual International Conference on Research and Development in Information Retrieval, in the press.

1. Execute a similarity search of a chemical database for some particular target structure using two, or more, different measures of inter-molecular structural similarity.
2. Note the rank position, r_i, of each database structure in the ranking resulting from use of the *i*-th similarity measure.
3. Combine the various rankings using one of the fusion rules (MIN, MAX or SUM), giving a new combined score for each database structure
4. Rank the resulting combined scores, and then use this ranking to calculate a quantitative measure of the effectiveness of the search for the chosen target structure.

Schema 1. Combination of similarity rankings using data fusion

Name	Fusion Rule
MIN	minimum (r_1, r_2 ,…r_i …r_n)
MAX	maximum (r_1, r_2,…r_i …r_n)
SUM	$\sum_{i=1}^{n} r_i$

Table 1. Fusion rules for combining *n* ranked lists, where r_i denotes the rank position of a specific database structure in the *i*-th ($1 \leq i \leq n$) ranked list.

Activity	2D	Phys	3D	MAX	MIN	SUM
L_1	1.40	3.05	1.12	2.96	2.02	3.25
L_2	2.14	3.50	5.93	4.36	3.57	5.00
L_3	5.53	6.35	3.69	6.00	5.81	6.16
L_4	5.33	4.44	4.06	5.17	4.67	5.5
M_1	2.29	6.17	2.50	4.96	4.08	5.04
M_2	6.48	5.00	5.52	6.52	5.86	6.17
N_1	2.71	4.43	2.71	3.19	3.29	3.81
N_2	3.67	4.19	3.67	4.00	4.24	4.00
Mean Actives	3.99	4.64	3.65	4.65	4.19	4.93
Mean Rank	4.63	2.88	5.00	2.94	3.50	2.01

Table 2. The mean number of actives in the top-10 rank positions for each activity class in the cellular-uptake dataset for the original similarity methods (columns 2D, Phys and 3D) and after data fusion (columns MAX, MIN and SUM). The shading indicates a fused result at least as good as the best original similarity measure for that target structure.

Target	3D	2D	FBSS	MAX	MIN	SUM
Apomorphine	15	23	14	24	16	26
Captopril	23	34	12	26	27	31
Cycliramine	43	31	36	43	42	45
Diazepam	27	27	15	23	23	22
Diethylstilb'ol	44	33	34	42	38	42
Fenoterol	19	33	17	28	29	31
Gaboxadol	6	2	6	5	6	5
Morphine	20	28	16	19	24	16
RS86	0	8	5	10	6	14
Serotonin	13	19	13	13	20	15
Mean Actives	21.0	23.8	16.8	23.3	23.1	24.7
Mean Rank	3.60	3.05	5.15	3.40	3.05	2.75

Table 4. The number of actives found in the top-50 rank positions for searches in the WDI database for the original similarity methods (columns 3D, 2D and FBSS) and after data fusion (columns MAX, MIN and SUM). The shading indicates a fused result at least as good as the best original similarity measure for that target structure.

	n=1			*n*=2			*n*=3			*n*=4			*n*=5		
Method	MAX	MIN	SUM	MAX	MIN	SUM	MAX	MIN	SUM	MAX	MIN	SUM	MAX	MIN	SUM
2D	<0.01	0.18	<0.05	<0.01	0.24	<0.01	<0.01	0.79	<0.01	<0.01	0.43	<0.01	<0.01	0.58	<0.01
3D	<0.01	0.84	<0.01	<0.01	0.53	<0.01	<0.01	0.43	<0.01	<0.01	0.43	<0.01	<0.01	0.42	<0.01
Phys	<0.01	0.34	<0.01	<0.01	0.58	<0.01	<0.01	0.54	<0.01	<0.01	0.33	<0.01	<0.01	0.38	<0.01

	n=6			*n*=7			*n*=8			*n*=9			*n*=10		
Method	MAX	MIN	SUM	MAX	MIN	SUM	MAX	MIN	SUM	MAX	MIN	SUM	MAX	MIN	SUM
2D	<0.01	0.42	<0.01	<0.01	0.32	<0.01	<0.01	0.38	<0.01	<0.01	<0.05	<0.01	<0.01	<0.01	<0.01
3D	<0.01	0.20	<0.01	<0.01	<0.05	<0.01	<0.01	<0.05	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01
Phys	0.43	0.60	<0.01	0.07	0.53	<0.01	0.11	0.43	<0.05	0.21	0.43	<0.05	0.36	0.28	<0.05

Table 3. The *p* values from the Wilcoxon test for rank positions *n*=1-10. Values ≤ 0.05 denote a fusion rule that is significantly better than an original similarity measure for the cellular-uptake dataset.

c=1		*c*=2		*c*=3		*c*=4		*c*=5		*c*=6	
2	0.80	23	1.10	23F	1.28	23FJ	1.52	23FJN	1.45	23FJNT	1.43
3	1.12	2F	1.04	23J	1.39	23FN	1.23	23FJT	1.69		
F	0.89	2J	1.01	23N	1.04	23FT	1.43	23FNT	1.36		
J	1.08	2N	0.68	23T	1.24	23JN	1.31	23JNT	1.43		
N	0.63	2T	0.95	2FJ	1.35	23JT	1.45	2FJNT	1.43		
T	0.69	3F	1.09	2FN	1.08	23NT	1.25	3FJNT	1.51		
		3J	1.25	2FT	1.28	2FJN	1.28				
		3N	1.00	2JN	1.03	2FJT	1.53				
		3T	1.32	2JT	1.10	2FNT	1.28				
		FJ	1.20	2NT	0.95	2JNT	1.17				
		FN	0.91	3FJ	1.40	3FJN	1.35				
		FT	1.11	3FN	1.19	3FJT	1.55				
		JN	0.89	3FT	1.33	3FNT	1.41				
		JT	0.93	3JN	1.25	3JNT	1.36				
		NT	0.85	3JT	1.45	FJNT	1.32				
				3NT	1.20						
				FJN	1.11						
				FJT	1.21						
				FNT	1.11						
				JNT	1.12						

Table 5. Mean number of actives found in the ten nearest neighbours when combining various numbers, *c*, of different similarity measures for searches of the Kahn dataset. The shading indicates a fused result at least as good as the best original similarity measure.

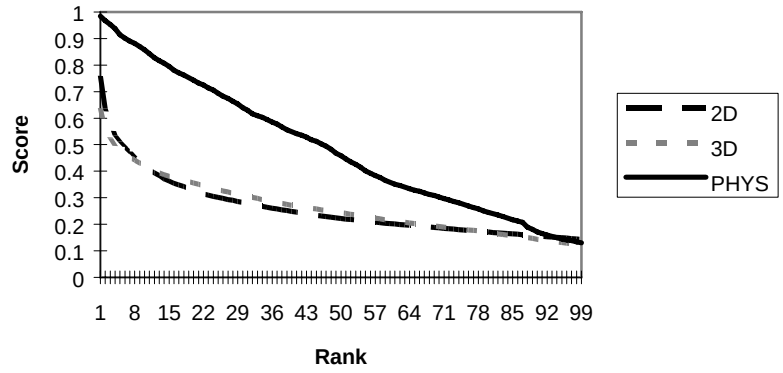

Figure 1. Plots of mean score against rank for the three types of original (*i.e.*, unfused) similarity measure for the cellular-uptake dataset.

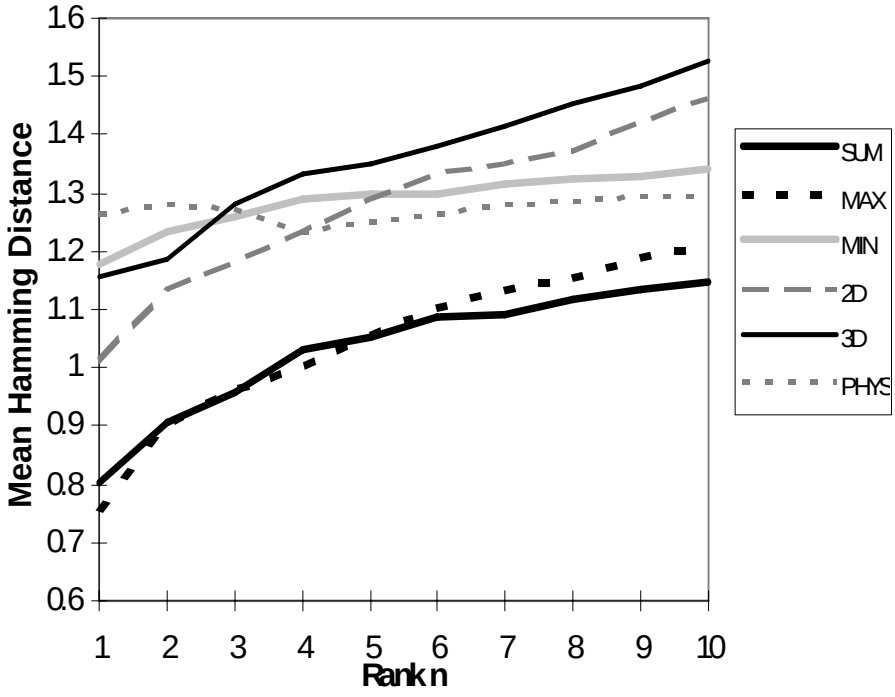

Figure 2. The mean Hamming Distance at each rank n, $1 \leq n \leq 10$.