J Comput Aided Mol Des (2010) 24:789–801
DOI 10.1007/s10822-010-9374-0

ElectroShape: fast molecular similarity calculations incorporating shape, chirality and electrostatics

M. Stuart Armstrong · Garrett M. Morris · Paul W. Finn · Raman Sharma · Loris Moretti · Richard I. Cooper · W. Graham Richards

Received: 5 January 2010 / Accepted: 30 June 2010 / Published online: 8 July 2010

Abstract We present ElectroShape, a novel ligand-based virtual screening method, that combines shape and electrostatic information into a single, unified framework. Building on the ultra-fast shape recognition (USR) approach for fast non-superpositional shape-based virtual screening, it extends the method by representing partial charge information as a fourth dimension. It also incorporates the chiral shape recognition (CSR) method, which distinguishes enantiomers. It has been validated using release 2 of the Directory of useful decoys (DUD), and shows a near doubling in enrichment ratio at 1% over USR and CSR, and improvements as measured by Receiver Operating Characteristic curves. These improvements persisted even after taking into account the chemotype redundancy in the sets of active ligands in DUD. During the course of its development, ElectroShape revealed a difference in the charge allocation of the DUD ligand and decoy sets, leading to several new versions of DUD being generated as a result. ElectroShape provides a significant addition to the family of ultra-fast ligand-based virtual screening methods, and its higher-dimensional shape recognition approach has great potential for extension and generalisation.

This research was financed in part by the 6th Framework Program of the European Union (DeZnIT, Project Number 037303) and by the National Foundation for Cancer Research (NFCR).

M. S. Armstrong (✉) · G. M. Morris · P. W. Finn · R. Sharma · R. I. Cooper
InhibOx, Pembroke House, 36-37 Pembroke Street, Oxford OX1 1BP, UK
e-mail: stuart.armstrong@inhibox.com

L. Moretti · W. G. Richards
Department of Chemistry, InhibOx Laboratory, University of Oxford, Pembroke House, 36-37 Pembroke Street, Oxford OX1 1BP, UK

Keywords Molecular similarity · Molecular descriptors · Ligand-based virtual screening · Drug design · Chirality chemotypes

Introduction

Ligand-based virtual screening has proven to be an important tool for identifying active compounds and discovering new drugs [1–3]. Since assembling experimental data is generally slow and expensive, preliminary screening by computers has been shown to provide a valuable and cost-effective alternative [4]. However, despite the continued growth in computational power, the individual pairwise comparisons are slow for many reported methods and a great many of these may need to be made, for example in a diversity analysis [5] or virtual screening against the very large databases available today. Thus, there remains a need for very fast ligand-based virtual screening methods.

Shape comparison methods are an important class of such ligand-based virtual screening approaches, as molecules which bind to the same biological target tend to have similar shape [6–13]. Ultrafast Shape Recognition (USR) [14, 15] is one of a class of very fast methods that seek to speed up the process by breaking the calculation up into two stages. First, a set of shape descriptors is computed for each molecule in the library to be screened. In the second step, molecules can be compared by computing the distance between these pre-computed shape descriptors, a process which is nearly instantaneous.

USR has proven to be effective in prospective virtual screening studies [16], but it has some limitations: it only encodes shape information, and does not distinguish between enantiomers. Previously, we reported Chiral Shape

Recognition (CSR), an extension of USR that calculates the descriptors in such a way as to distinguish enantiomers [17] (see Sect. 2.2). Here, we further extend USR/CSR-like shape recognition methods to include electrostatic complementarity. This is a significant enhancement, since electrostatic complementarity is a key aspect of molecular recognition that is neglected by purely shape-based approaches. We call this generalisation ElectroShape, since it combines electrostatic (partial charge) information and shape within a single approach.

The essence of ElectroShape is based on the realisation that atomic coordinates are simply numbers, and that the partial charge is just another number. Hence the partial charge defines a fourth coordinate, with atoms being identified by points in four-dimensional space. The mathematical framework is then applied in four dimensions rather than three, to generate the combined descriptors. Searching can then be performed in an analogous manner to USR, while retaining all the speed advantages of that approach.

We believe this to be the first time that electrostatic information has been combined with a fast shape-based virtual screening method. It is not only innovative, but effective, nearly doubling the average enrichment ratio at 1% as compared with USR and CSR (see Sect. 4), a result statistically significant at the 99% level. The value of this approach is further established by the fact that electrostatic information, *on its own*, gave a higher enrichment result than did shape-based USR (see Sect. 4.1). Hence the inclusion of partial charge is an important step forward in the field of non-superpositional ligand-based screening, and opens up many avenues for future improvement. Why electrostatic similarity alone should have such a good performance is an interesting open question that warrants further study.

Validation studies of ElectroShape were initially performed using the current release (release 2) of the Directory of Useful Decoys [18], which has been specifically designed to be a stringent test of virtual screening methods. During the course of these studies, we obtained unexpected results which cast doubt on the consistency of the partial charge values for the DUD decoys. This has been confirmed by the DUD distributors—a new release is planned shortly removing those errors. Therefore we calculated alternative partial charge assignments using AMSOL [19, 20], Gasteiger [21], AM1 [22] and the modified MMFF94 [23] implemented in CCG MOE [24] (the difference between standard MMFF94 and MOE's MMFF94x is that in the later, hydrogen atoms on alkane carbons have a non-zero partial charge). We used these new DUD sets in our validation studies, and they have been made available online [25].

To demonstrate that ElectroShape is not simply a sophisticated substructure search method only capable of finding actives that share a common chemotype with the query ligand, further validation was carried out by taking into account chemotype redundancy in the DUD active sets as described by Cheeseright et al. [26]. The results demonstrate that ElectroShape has good cross-chemotype enrichments.

Importantly, through the developments described herein, we have developed in ElectroShape a platform that has great capacity for generalisation. The example presented here adds partial charge as a fourth dimension, but any relevant atom-centred information could be used, such as atomic lipophilicity, atomic solvation parameters, and even the electric field at the atom *etc*. Its high performance, combined with its capacity for generalisation, makes it an excellent tool for virtual screening.

ElectroShape recognition

The primacy of shape complementarity for molecular recognition has long been recognised and many methods have been described for molecular shape comparison [6]. However, it is well-known that electrostatic complementarity is also very important, and so a method that combines electrostatic and shape similarity is desirable. As mentioned in the introduction, ElectroShape achieves this by considering the partial atomic charges in addition to the spatial x, y and z positional coordinates. Each atom can now be described by four numbers: the three Cartesian coordinates and the partial charge, and, as such, can be seen as a point (x, y, z, q) in the four dimensional Euclidean space $\mathbb{R}^4$. In this manner, molecules can be compared using any shape recognition program if it is extended to consider their shapes in $\mathbb{R}^4$. For example, consider the two one-dimensional molecules carbon dioxide and hydrogen cyanide (Fig. 1). These molecules are very similar in terms of steric shape, but look very different when the partial charges are added as an extra dimension.

The distance function in $\mathbb{R}^4$ is similar to that in standard three dimensional space: if $p = (x, y, z, q)$ and $p' = (x', y', z', q')$, then the distance between p and p' is given by:

$$||p - p'|| = \sqrt{(x - x')^2 + (y - y')^2 + (z - z')^2 + (q - q')^2}$$

One shape-recognition method which scales well to four-dimensional space is USR [14, 15]. In USR, four points, called "centroids", are identified based on the three dimensional structure of the molecule. From each of these centroids, the distances to every atom in the molecule are computed, generating four distance distributions. Dimensional considerations imply that these distributions generically encode all of the shape information. The first three moments of each distribution are extracted and normalised, giving a vector of twelve numbers. Molecules can then be compared rapidly using these twelve descriptors.

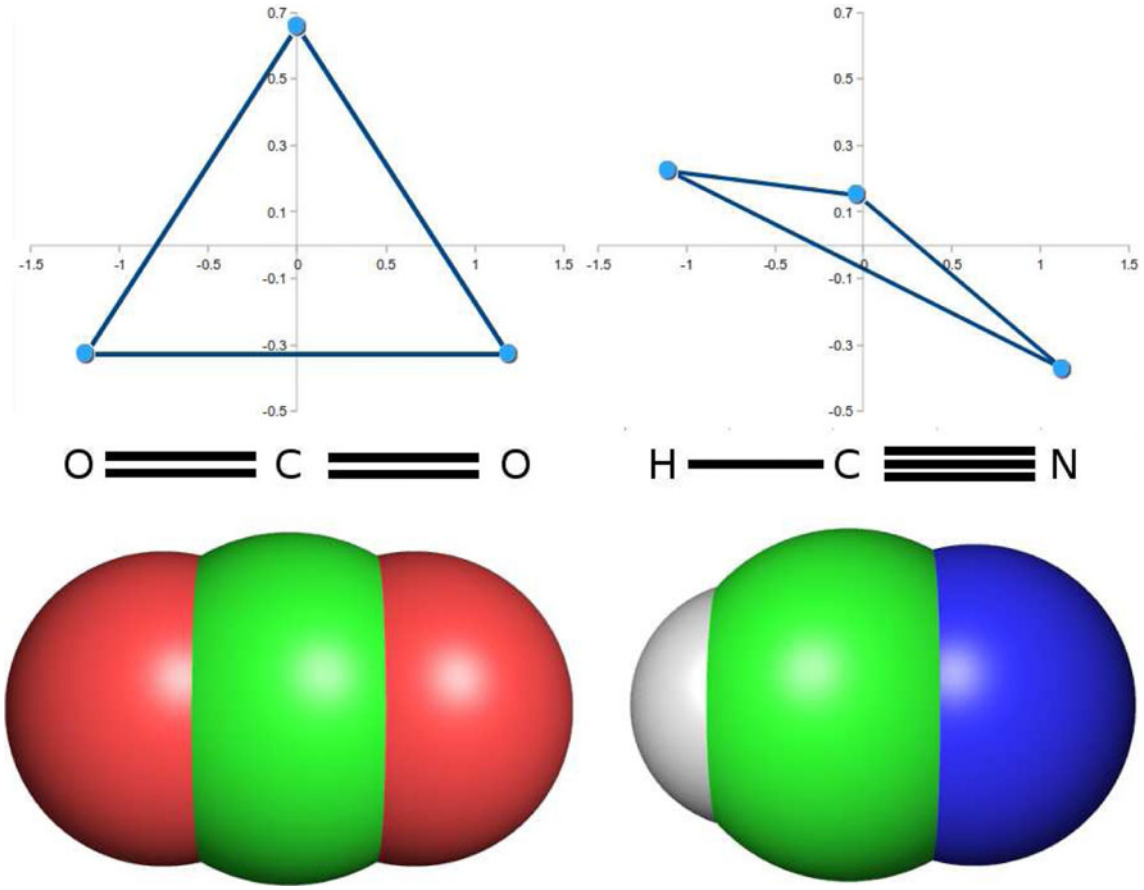

Fig. 1 Adding partial charge as an extra dimension helps to discriminate between similarly-shaped molecules. The partial charges and geometries were computed using AMSOL [19, 20], with options AM1, CM2 and geometry optimisation with a maximum of 10,000 iterations

There is no inherent reason to limit the method to four dimensions, nor to limit the extra quantity to being purely the partial charge. Any other numerical information could be used, in a natural and straightforward generalisation to higher dimensions – as long as the numerical data has a cardinality that is chemically meaningful.

There remains one last subtlety: the partial charge and the spatial positions are not given in the same units, and there is no natural way of converting electron charges into Ångstroms. This issue will be addressed in Sects. 2.3 and 4.1.

Distance distributions and moments

In general, to find the exact position of a point in the n-dimensional space $\mathbb{R}^n$, the distances from $n + 1$ fixed 'centroids' are needed (see Fig. 2, where three centroids are not enough to decide between the two white dots). To avoid degeneracy, these centroids must be as linearly independent as possible; in other words, the $n + 1$ centroids must not lie in an $n - 1$ dimensional subspace. Thus, with five centroids fixed in four dimensional space $\mathbb{R}^4$, it is possible to unambiguously determine the position of the (x, y, z, q) vector of all the atoms in a molecule from the calculated distances from each centroid to each atom.

The centroids themselves are defined in terms of the atom positions; indeed the positions of the centroids are not needed to determine the shape of the molecule: the five distributions are enough to define it in nearly all cases (as long as there are more atoms than centroids). This can perhaps best be seen from the perspective of information theory: a generic collection of n points in $\mathbb{R}^4$, plus five centroids, will be defined by $4(n + 5)$ numbers. Ignoring the effects of translations and rotations, there are $4(n + 5) - 4 - 10$ independent variables here (though the centroid positions are actually dependent variables). The distributions encode $5n$ numbers; so it is not surprising that for large n, we can extract the information in the first scenario from the second.

Replacing the $4n$ coordinates of the n atoms with $5n$ distance measures does not simplify the problem. However, defining positions in terms of distributions has two clear advantages: firstly, they are already independent of translations and rotations, and secondly it is easy to summarise distributions with simpler information. For each

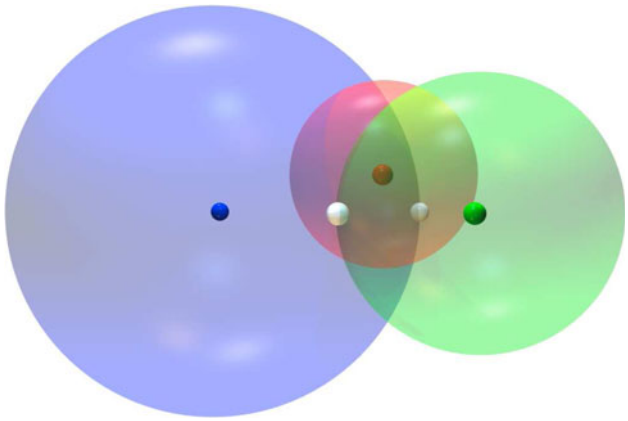

Fig. 2 Three spheres around each centroid intersect at two points (shown in *white*). To unambiguously determine the position of a point in three dimensions, a fourth centroid is required

distribution, we extract three numbers: the mean, the standard deviation, and the cube root of the third central moment. This third measure is given, for a distribution $\{x_1,..., x_n\}$ with mean m, by the following formula:

$$M_3 = \sqrt[3]{\sum_{i=1}^{n} \frac{(x_i - m)^3}{n}}.$$

The advantage of these three moments is that they all have the same units, unlike the skewness, for instance, which is unitless. Thus each molecule is assigned fifteen numbers, three numbers for each of the five distributions, making a 15-dimensional vector. The process of calculating these fifteen numbers is very quick, and once they are assigned, molecules can be compared by calculating the distance between their 15-vectors of moments.

The choice of the centroid positions

The choice of the method for positioning the five centroids was dictated by several criteria:

- Ease of computation: the process of determining the centroids must be simple and easily understood.
- Continuity: small changes to the molecule should result in small changes in the position of the centroids.
- The positions of the centroids must distinguish a chiral molecule from its enantiomer.
- Each centroid must generate a large amount of useful information; consequently, each centroid must be neither too close to another centroid, nor too far away from the molecule as a whole.
- Avoidance of degeneracy: the centroids, if feasible, must be as linearly independent as possible: no three centroids should be on a line and no four centroids should be coplanar.
- Invariance: the choice of the centroids must not depend on the choice of coordinates for the $\mathbb{R}^3$ component.

Unfortunately, it is not possible to satisfy all these criteria; satisfying continuity and ease of computation simultaneously is particularly tricky—the reason being that concepts such as 'closest atom' or 'furthest atom' are *not* continuous concepts (e.g. small changes in conformations can make the identification of the furthest atom jump discontinuously). The correct way of rendering continuous the idea of 'furthest atom', say, is to take a weighted average of all atoms that gives high weight to large distances. This is computationally and conceptually much more complicated, would give the same result for molecules that are not close to the zones of discontinuity, and, in some cases, might result in not being able to assign a separate centroid. For instance, any weighted mean of the atoms from the centre of mass of a buckminsterfullerene molecule would be again the centre of mass; an arbitrary (and discontinuous) choice of the furthest atom from the centre of mass would allow encoding of much more information about the shape.

Thus the requirement of continuity was dropped, but all other criteria were satisfied. In practice, very few molecules will be close to the zones of discontinuity, so not much is lost by this choice.

Distinguishing chiral molecules is particularly important, and a method to achieve this with a suitable choice of centroids has recently been published [17]: this is the Chiral Shape Recognition (CSR) method, which is most properly seen as a three-dimensional version of ElectroShape.

For ElectroShape, the first centroid, c_1, is defined as the unweighted barycentre of all atom positions in $\mathbb{R}^4$ (the geometric centre of the molecule). The second centroid, c_2, is defined as the atom position furthest from c_1 in $\mathbb{R}^4$ (see Sect. 2.3 on converting charge to distance), and the third centroid, c_3, as the atom position furthest from c_2. These are three of the four centroids defined for USR; the remaining USR centroid (the atom closest to c_1) is not used for ElectroShape.

These three centroids give a pair of vectors $\vec{a} = c_2 - c_1$ and $\vec{b} = c_3 - c_1$. The three-dimensional spatial parts of these vectors—the first three coordinates—are designated by $\vec{a}_S$ and $\vec{b}_S$.

It is now possible to define the three dimensional vector $\vec{a}_S \times \vec{b}_S$. Since the cross product is *not* equivariant under reflections, this vector distinguishes between a chiral molecule and its enantiomer. To keep the units consistent and to satisfy the other criteria, this vector is scaled to be half the length of vector $\vec{a}$:

$$\vec{c} = \left(\frac{||\vec{a}||}{2||\vec{a}_S \times \vec{b}_S||} \right) \vec{a}_S \times \vec{b}_S$$

This purely spatial vector was then added to c_1 to define the spatial component of the fourth and fifth centroids. Their charge component was assigned by simply taking the highest and lowest partial atomic charges of the molecule:

thus if q_+ is the highest partial charge, and q_- is the lowest, the remaining two centroids are:

$$c_4 = (c_1)_S + \overrightarrow{c} + (0, 0, 0, \mu q_+)$$
$$c_5 = (c_1)_S + \overrightarrow{c} + (0, 0, 0, \mu q_-),$$

where $(c_1)_S$ is the spatial component of c_1, and μ is a scaling factor converting electron charges to Ångstroms (see next Section).

A question of scale

It is important to ensure that the fourth dimension is scaled consistently with respect to the spatial coordinates, which poses the question: how many Ångstroms are there to one electron charge? ElectroShape is thus dependent on a scaling factor μ, so that an atom with positional coordinates x, y and z in Ångstroms and partial charge q is described by coordinates in four dimensional space as:

$(x, y, z, \mu q)$.

The unit of μ is Ångstrom per electron charge. By choosing a very small μ, the method becomes a purely shape-based recognition program, as the charge contribution becomes insignificant; conversely, choosing a very large μ results in a comparison purely on the basis of partial charges, as the spatial coordinates' contribution becomes negligible.

The optimal ratio was experimentally determined by choosing the value that gave the best average enrichment in our validation studies using the DUD dataset (see Sect. 4.1).

Similarity score

Each molecule is assigned a vector of centroid distance distribution moments consisting of fifteen numbers. It is necessary to compute a measure of the distance between molecules, so that they can be compared and ranked. The Manhattan distance provides one such measure, so-called because it is the generalisation of the distance function for taxis driving on a grid of perpendicular roads, such as in Manhattan. For the moment vectors $\overrightarrow{m} = (m_1, m_2, \ldots, m_{15})$ and $\overrightarrow{n} = (n_1, n_2, \ldots, n_{15})$, the Manhattan distance between the two is:

$$d_{\mathrm{Man}}(\overrightarrow{m}, \overrightarrow{n}) = \sum_{i=1}^{15} |m_i - n_i|.$$

Following convention, this distance is normalised by dividing by the number of descriptor values, which in this instance is 15. Since it is an unbounded distance metric, it is not entirely suitable; however, a simple inversion changes this distance measure into a value that varies between zero and one, with one corresponding to complete equality and the unreachable zero corresponding to infinite distance between the two molecules in this scale. This is the right behaviour for a similarity score, and the formula becomes:

$$d_{\mathrm{InvMan}}(\overrightarrow{m}, \overrightarrow{n}) = \frac{1}{1 + \frac{1}{15} d_{\mathrm{Man}}(\overrightarrow{m}, \overrightarrow{n})} = \frac{1}{1 + \frac{1}{15}\left(\sum_{i=1}^{15} |m_i - n_i|\right)}.$$

The inverse Manhattan distance between the ElectroShape descriptors of two molecules is called the ElectroShape similarity score.

Number of free parameters in the method

ElectroShape has one free parameter, μ, the ratio of distance to charge. During the initial testing stages, the Euclidean distance metric was occasionally replaced with a mixed Euclidean–Manhattan metric distance metric; as this resulted in slightly poorer performance, this metric was rejected. Four ways of assigning the centroids were considered, before settling on the one here. An extension of the ElectroShape descriptors was also investigated, in which a fourth moment was computed for each centroid distance distribution (see Sect. 4.3).

The construction of the method is thus dependent on one free parameter and four binary choices.

Testing methodology

The DUD datasets

To test the enrichment characteristics of the method, ElectroShape was validated using release 2 of the DUD dataset [18], a collection of datasets for forty diverse targets, ranging from ACE (angiotensin-converting enzyme) to VEGFR2 (vascular endothelial growth factor receptor kinase). For each target, DUD supplies two sets of molecules: one containing known active ligands for the target, and one containing decoy molecules. The decoys are presumed not to bind with the given target, but have similar molecular properties to the actives for that target (charge, chemical formula, molecular weight, etc.). These sets were assembled so that any screening program differentiating actives from decoys could not do so based on simple molecular properties alone. The DUD release 2 molecules' partial charges were assigned using AMSOL; for more details on the conformer generation and the partial charge assignments, see [18].

These datasets were used as supplied for enrichment studies, at 1% of the database screened, and for constructing ROC curves (see the next section). Various alternate partial charge calculation methods were also used (in one case, with geometric optimisation); the details of these are in Sect. 3.3.

Virtual screening experiments using ElectroShape were run for each of the forty targets' datasets. Each active in turn, in each target set, was taken as the query molecule, and the ElectroShape similarity score from that active to every other active and decoy in that same target set was then computed. A good performance would correspond to actives being found to be more similar to each other than to the decoys.

Performance measures

The performance of the ElectroShape method was analysed in three ways: by computing the AUC (area under the curve) for the rank ROC (receiver operating characteristic) curves and the similarity ROC curves, and the enrichment ratios at 1% of the database screened.

ROC curves

ROC curves plot the proportion of true positives relative to the proportion of false positives, as the detection threshold of the system is varied. To generate the ElectroShape ROC curves, the ElectroShape similarity between the actives and all molecules are first calculated using the inverse Manhattan distance. Then each active A_i is treated as query molecule in turn. This generates, for each given active A_i, an ordered list of all the other actives and decoys in terms of their ElectroShape similarity score to A_i. If there are n actives and m decoys, this results in n lists with $m + n - 1$ molecules in them, because, in each case, the query molecule itself is excluded. Hence there is a total of $n(n - 1)$ actives across the lists (true positives) and nm decoys across the lists (false positives).

The ROC curves are computed from these lists. The rank ROC curve uses the rank on the lists as the varying threshold, while the similarity ROC curve uses the similarity score as the varying threshold. The difference between the two methods of combining the ranked ligands and decoys can best be illustrated by example. Assume there are two actives, A_1 and A_2, and a single decoy D, with the similarity scores as presented in Table 1. The rank ROC curve starts at (0, 0), for the 'zeroth' rank. When the rank threshold is reaches one, one active and one decoy will be found—50% of the total in each case. Hence the next point on the ROC curve is (0.5, 0.5). Finally, when the rank threshold is set to two, two actives and two decoys will have been found—all of them. Hence the last point on the ROC curve is (1, 1).

Table 1 Example of ordered lists from different actives

A_1			A_2		
Molecule	Similarity	Rank	Molecule	Similarity	Rank
A_2	0.8	1	D	0.9	1
D	0.5	2	A_1	0.8	2

The similarity ROC curve is different, however. It starts at (0, 0) for a similarity threshold of 1. When the similarity threshold is lowered below 0.9, one decoy will be found, 50% of the total: so the next point on the curve will be (0.5, 0). When the similarity threshold is lowered below 0.8, both actives will be found, so the next point on the curve is at (0.5, 1). Finally, once the similarity threshold is lowered below 0.5, the last decoy is found, and the last point on the ROC curve will be (1, 1).

In this simple example, though the rank ROC curve is evidently better for early enrichment, the areas under the curve (AUC) for both are the same. With the addition of more decoys and actives, the AUC's can diverge quite considerably; in general, there is no a priori way of knowing which method will be the best. Methods which produce good similarity ROC curves are those for which the similarity score is a useful intrinsic measure; however those that produce good rank ROC curves are those for which the ranking generated by the similarity score is the useful measure. This affects how the method should be used.

The similarity score is un-normalised (there is no factor in it to compensate for the sizes of the molecules), and hence, mathematically, our hypothesis is that the rank ROC curve will demonstrate a better performance than the similarity ROC. As shall be seen in Sect. 4, ElectroShape indeed is a method with better rank ROC results than similarity ROC results, and this effect is most strongly visible for early enrichment.

Enrichment

The enrichment ratio at $n\%$ of the library screened is a simple numerical measure, hereafter designated by $E_{n\%}$. It is calculated by looking at the top $n\%$ closest molecules to a given active. The enrichment is the ratio of the proportion of actives in this closest subset to the proportion of actives that would be expected by chance alone. The enrichment is averaged across the active set, to get the $E_{n\ \%}$ for the entire target set. Typically, this result will be further averaged across the 40 DUD datasets.

An important property of enrichment ratios is that for an active set that represents x % of the total set, $E_{n\%}$'s maximal value is $\min\{100/n, 100/x\}$. For pure chance, $(x \times n / 100)\%$ of the total database will consist of actives among the first $n\%$ closest molecules. However, the maximum number of actives in the first $n\%$ closest molecules is $\min\{x\%, n\%\}$ of the whole database (since there cannot be more actives than actives exist, or more than the size of the subset being looked at), giving the result.

We will choose $E_{1\%}$ to simulate a typical virtual screening experiment. Though the DUD website [27] indicated that

there are 36 decoys for each active, this number varies from target to target. COMT, for instance, has 11 actives and 468 decoys, hence the ligands make up 2.30% of the combined set. This is the lowest percentage across the 40 targets; the highest is p38, with 4.73%. The enrichment experiments performed in this paper involve using one active molecule as a query, which is used to rank the other actives and decoys; hence there is one fewer active in each enrichment experiment. Once this is taken into account, the percentage of actives in COMT is 2.09 %, while that in p38 is 4.72%. Hence the maximum possible $E_{1\%}$ varies from 47.8 for COMT to 21.2 for p38, with an average maximum of 36.4; these maximum values are plotted in Fig. 8.

Problems with the DUD datasets

Initial studies on the standard release 2 DUD datasets revealed a problem. For a distance to charge ratio μ of 25 Å per electron charge, the average enrichment ratio at 1% is 30.2—while the maximum theoretical enrichment ratio at 1%, based on the relative sizes of the active and decoy sets, is 36.4. However when μ was varied between 25×2^{-10} and 25×2^{10}, the average enrichment ratio at 1% changed as shown in Fig. 3 (the enrichment ratio at 1% of two of the datasets—EGFR and GART—are also plotted for illustrative purposes).

The extreme right of this graph has a μ of 2.56×10^4 Å per electron charge—implying that the spatial arrangement of the molecule is irrelevant. Furthermore, the performance does not dip to any significant extent as μ is increased. This behaviour was very peculiar, as it suggested shape was irrelevant to the enrichment score.

To investigate this further, a detailed analysis of the DUD charges was performed. In order to have a basis for comparison, the datasets were re-processed, reassigning the partial charges using AMSOL with AM1, CM2 and geometry optimisation—this being the closest we were able to come to implementing the method reported in the original DUD publication [18].

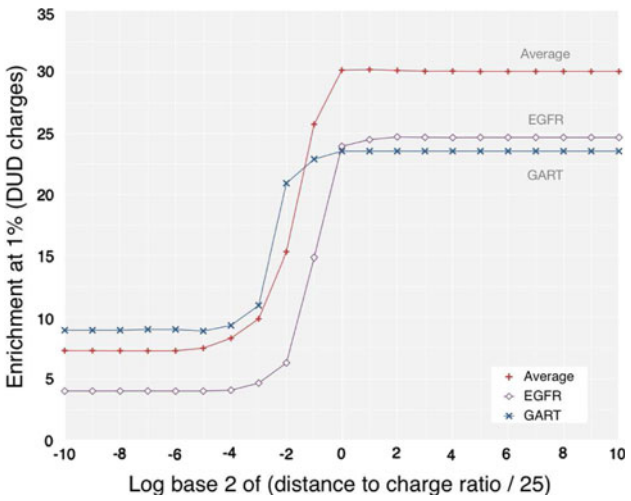

Fig. 3 Enrichment ratios at 1% for the release 2 DUD datasets and varying μ, the distance to charge ratio

For each molecule in the original DUD (release 2) sets, the standard deviation of the partial atomic charges for each molecule was calculated, giving a single measure of 'spread of charges' for each molecule. The average of this measure was then calculated for all the molecules in each active set in DUD, and again for each set of decoy molecules, giving 80 different numbers. The same was done for the recalculated AMSOL optimised datasets.

This average, for each ligand and decoy set from DUD release 2, was compared with its counterpart in the recalculated version. As Fig. 4 demonstrates, there is little difference in this average measure between the ligand sets, implying that our methods of charge assignment differ little from that used in the DUD original. However, the differences for the decoy sets are large (Fig. 5). This strongly suggests that the partial charge assignments for the DUD decoy sets are suspect. This was confirmed by the authors of the DUD datasets (personal communication) and a new revision of DUD will be made available online shortly.

This stimulated us to extend the study to explore several partial charge calculation methods and asses their effects on performance. Initial experiments showed that there were no problems with the conformations of the DUD datasets—we generated alternate conformers using MOE (MMFF94x forcefield), and substituted back in the original DUD charges, without any notable change in performance.

when the conformations were recalculated by MOE's conformer generation using the MMFF94x force-field (leaving the charges as they were), the enrichment results were essentially the same.

Thus, in addition to the optimised AMSOL charges, Gasteiger charges, AM1 charges and the MMFF94x charges were assigned using CCG's MOE, without any conformational optimisation. We report the results with the original DUD charges as well, for comparison with other electrostatic-based methods that may have used the original release 2 DUD datasets. For comparison with ElectroShape, the results for USR and CSR are also included.

The re-processed DUD datasets with their alternative charge assignments presented here can be downloaded from the InhibOx website [25].

Results

Figures 6 and 7 show the averaged rank ROC curves and similarity ROC curves respectively, for USR, CSR and ElectroShape with the five different methods of charge assignment.

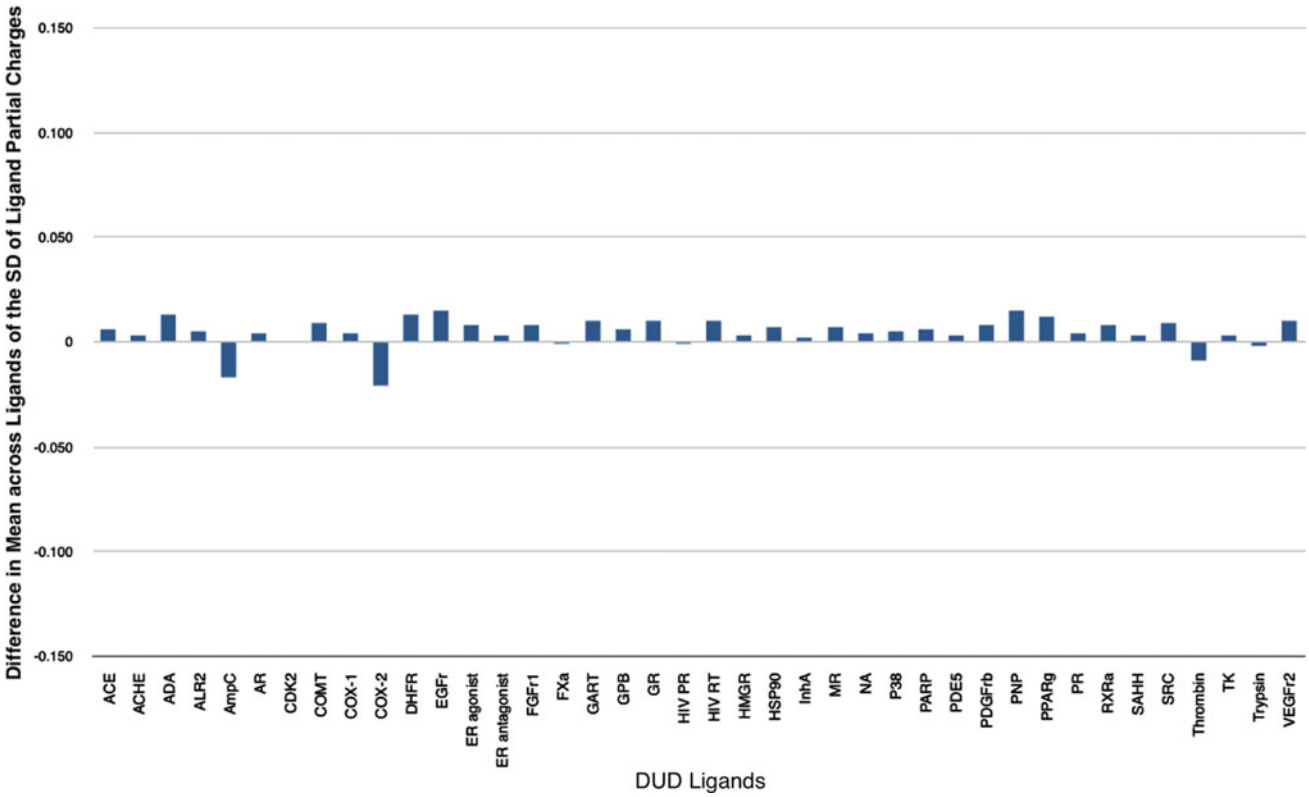

Fig. 4 The difference in mean of the standard deviation of each ligand's partial charges across all the active ligands in each DUD target set, between the original DUD release 2 partial charges and the AMSOL optimised charges computed for this study

Figure 8 breaks down the enrichment results for each of the forty datasets, for USR, CSR, ElectroShape with CCG MOE MMFF94x charges, and ElectroShape with the original DUD release 2 charges (for reference). The theoretical maximum enrichment at 1% is also plotted, demonstrating how close ElectroShape approaches this maximum with the original DUD charges.

Finally, Table 2 presents the averaged enrichment ratios at 1 % of the database screened, along with the averaged area under the curve (AUC) for the different ROC curves of the various methods (the standard deviation of these averages appear in brackets). Note that the figures for USR and CSR are not the same as reported in [17], where the DUD datasets were deliberately seeded with enantiomers. Also note that although the AUC's are different between the two methods of generating ROC curves; this difference is generally only at the third significant figure, which it would be spuriously precise to present here.

ElectroShape demonstrates a clear improvement in performance over USR and CSR in enrichment studies using the DUD datasets, for all four partial charge methods presented here (and for the original DUD charges). The $E_{1\%}$ values in Table 2 are notable, with MMFF94x demonstrating a near-doubling in enrichment ratios over the purely spatial methods.

Statistically, for a sample size of 40 datasets and a dependent paired one-sided t-test, CSR performs better than USR for $E_{1\%}$ at the 95 % significance level. Similarly, all versions of ElectroShape have a significantly better $E_{1\%}$ performance than USR and CSR at this level. At the 99% level, all versions of ElectroShape still outperform USR and CSR, but there is no longer any statistically significant difference between those two.

In terms of area under the ROC curves, the improvements are smaller, but the early parts (false positive rates <0.2) of Figs. 6 and 7 demonstrate that ElectroShape has an excellent early enrichment: all the AUC gains come early in the ROC curve, making ElectroShape an excellent initial filter.

Statistically, for rank AUC, there is no significant difference between USR and CSR. ElectroShape is significantly better than USR and CSR (at the 95% level) only for AMSOL OPT, MMFF94x and the original DUD charges. At the 99% level, only the improvement for the original DUD charges remains significant. The situation is nearly identical for similarity AUC, except that ElectroShape with MMFF94x now fails to be significantly better, at the 95% level, than USR or CRS.

The rank ROC curves (Fig. 6) perform better than the similarity ROC curves (Fig. 7) in the early part of the curve; for instance, the area under the MMFF94x curve for

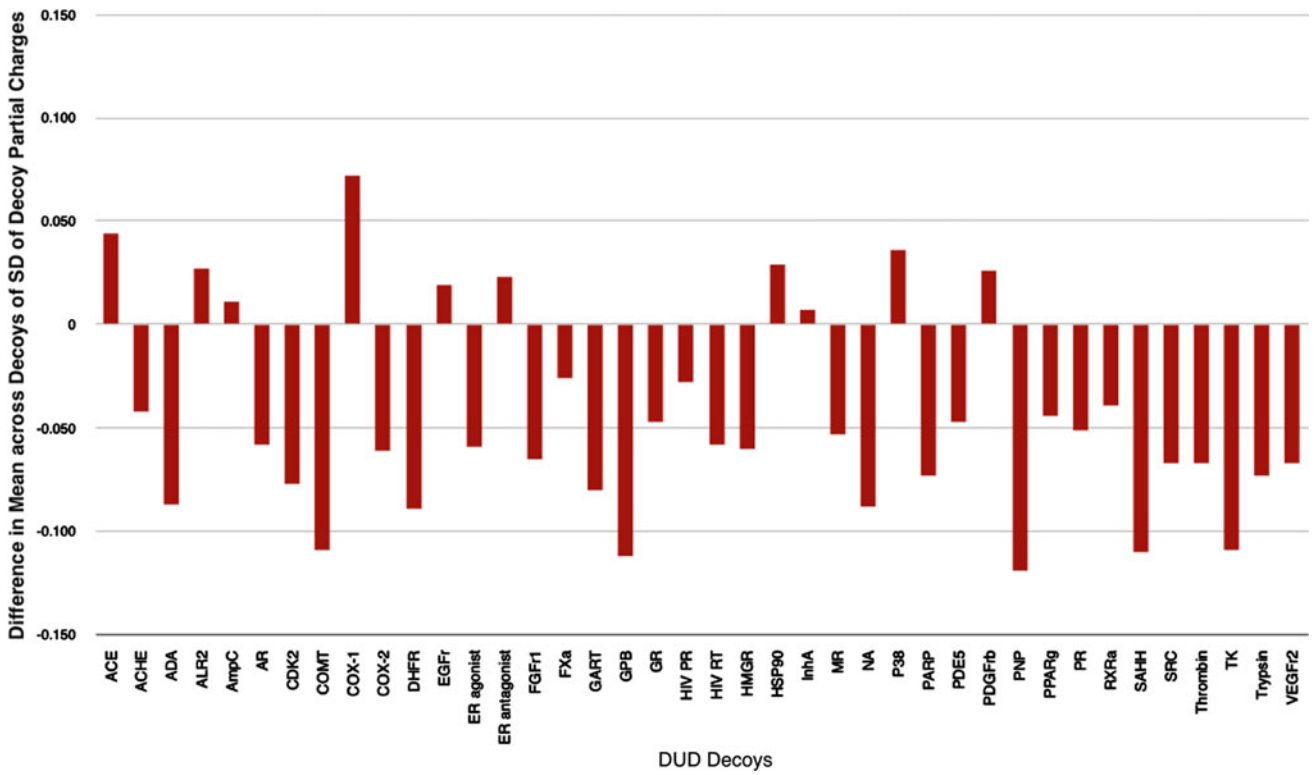

Fig. 5 The difference in mean of the standard deviation of each decoy's partial charges across all the decoys in each DUD target set, between the original DUD release 2 partial charges and the AMSOL optimised charges computed for this study (the *y*-axis scale is the same as in Fig. 4)

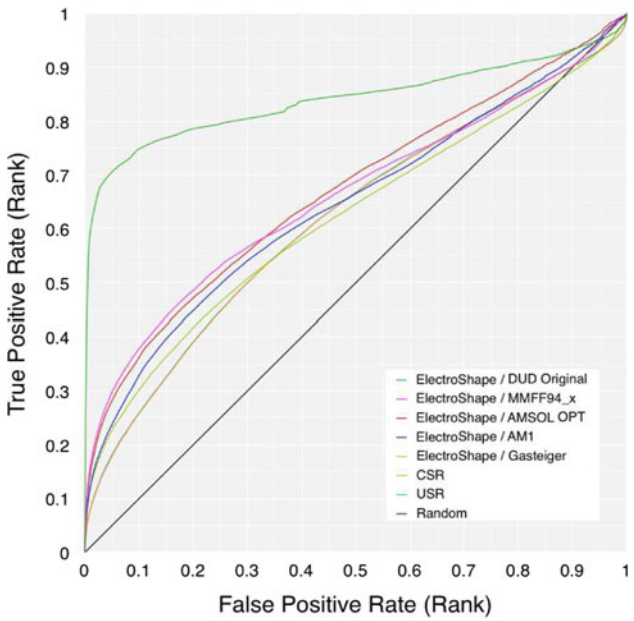

Fig. 6 Rank ROC curves for USR, CSR and ElectroShape with DUD release 2, MMFF94x, optimised AMSOL, AM1 and Gasteiger charges (at a ratio of 25 Å per electron charge, see Sect. 4.1)

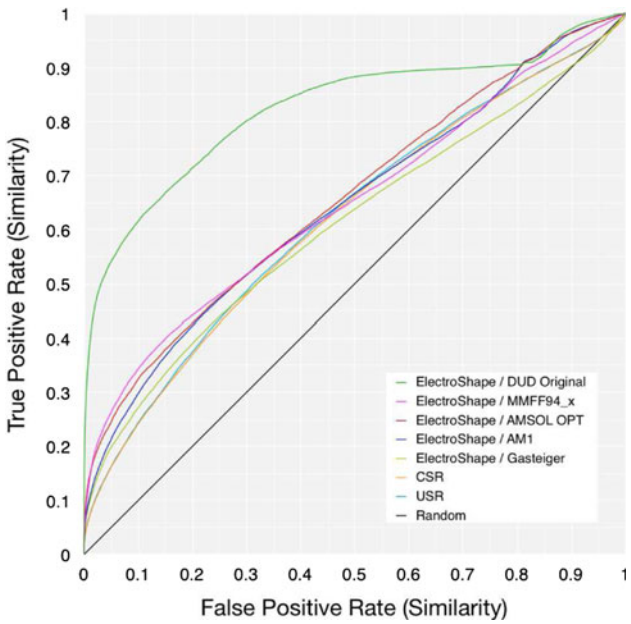

Fig. 7 Similarity ROC curves for USR, CSR and ElectroShape with DUD release 2, MMFF94x, optimised AMSOL, AM1 and Gasteiger charges (at a ratio of 25 Å per electron charge, see Sect. 4.1)

y values between 0 and 0.1 is 0.027 for the rank ROC curve, and 0.025 for the similarity ROC curve, and the relative difference increases as the interval decreases.

Statistically, choosing MMFF94x for the comparison (the best performer among our charge assignment methods), and doing a dependent paired one-sided t-test, the

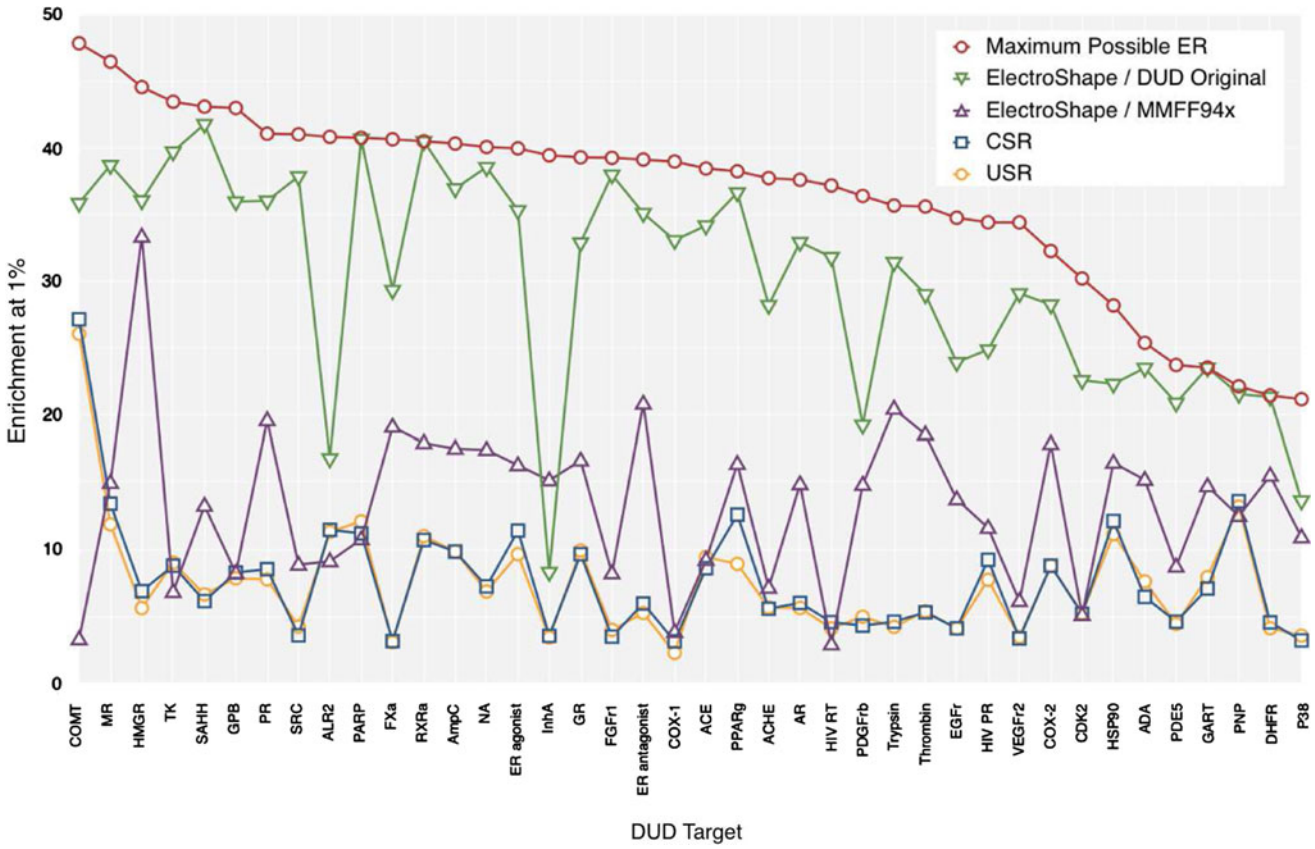

Fig. 8 Enrichment ratios at 1% across different target sets, for USR, CSR, ElectroShape using MMFF94x charges from CCG MOE, and ElectroShape using the original charges from DUD release 2 (at a ratio of 25 Å per electron charge, see Sect. 4.1)

Table 2 Summary of results for USR, CSR and ElectroShape (at a ratio of 25 Å per electron charge, see Sect. 4.1), averaged across the entire DUD release 2 datasets, with standard deviations in brackets

Method	Charges	$E_{1\%}$	Rank AUC	Similarity AUC
USR	–	7.4 (±4.2)	0.62 (±0.11)	0.63 (±0.10)
CSR	–	7.7 (±4.5)	0.62 (±0.11)	0.63 (±0.10)
ElectroShape	AM1	10.8 (±5.2)	0.64 (±0.14)	0.64 (±0.13)
	Gasteiger	10.8 (±5.3)	0.62 (±0.14)	0.62 (±0.13)
	AMSOL OPT	12.6 (±6.5)	0.67 (±0.15)	0.67 (±0.14)
	MMFF94x	13.3 (±6.0)	0.66 (±0.15)	0.66 (±0.14)
	DUD original	30.2 (±8.2)	0.84 (±0.14)	0.83 (±0.14)

rank AUC is statistically better, at the 95% level, than the similarity AUC. A one-sided test is justified in this case, as we formed the hypothesis that ranking was better than similarity based on mathematical intuition (see Sect. 3.2.1). This implies that ElectroShape works better as a ranking mechanism than as a similarity score calculator.

Finally, it should be noted that the original DUD release 2 charges give unusually good results by all measures. Though the issues with the DUD release 2 charges imply that this is not a fair assessment of ElectroShape's potential, we have included these results for comparison with other enrichment methods that used the charges of the original DUD datasets.

Shape and charge scaling

By changing the value of μ, the ratio of distance (in Å) to charge (in electron charges), it is possible to vary the importance of the spatial position relative to partial charge. How does enrichment performance vary with μ?

The value of μ was tested between 25×2^{-10} and 25×2^{10}, in increasing powers of two. The average enrichment ratios at 1%, using the original charges from DUD release 2, were already presented in Fig. 3; similar plots for the other charge assignments are given in Fig. 9.

As can be seen, the best performing alternative charge calculation, MMFF94x and AMSOL OPT, peak at $\mu = 25$,

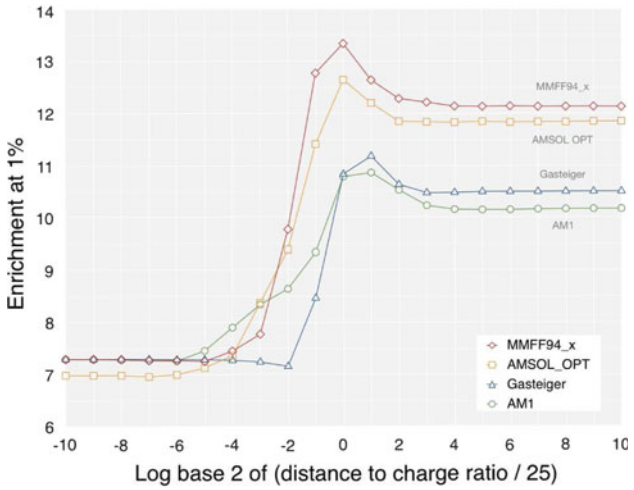

Fig. 9 Enrichment ratios at 1% for alternative charge assignments and varying μ, the distance to charge ratio, averaged across the 40 DUD datasets

while AM1 and Gasteiger peak at $\mu = 50$. Since the best average enrichment ratio at 1% was found using the MMFF94x charges from MOE, it was decided to use $\mu = 25$ as the default value for ElectroShape.

Most interestingly, "electrostatic only" performance is better than "shape only" performance. Why this should be so is an interesting open question that should be studied further. Howerver, as it stands, this observation can be seen as a vindication of the importance of partial charge in enrichment studies. It does strongly hint that any non-superpositional ligand-based enrichment method that does not make use of partial charge is forgoing a very important element, and could be improved by its inclusion.

Chemotypes

Good and Oprea [28] have shown that the active sets for each target in DUD often share similar chemotypes, and thus any virtual screening method should demonstrate that it is doing more than simply rediscovering actives with the same chemotype. Following the approach of Cheeseright et al. [26], we generated a reduced version of DUD for 13 of the targets, with chemotype information added, using the ZINC codes included in their supplementary information. We used the DUD sets with MMFF94x charges for this purpose.

To provide a baseline for comparison, the ElectroShape method was first run on the thirteen reduced datasets without using the chemotype information. This gave an average enrichment ratio at 1% of 11.1. Then the enrichment scores were calculated taking the chemotype information into account. Several different ways of doing this are suggested in the literature, but none is readily interpretable as an enrichment ratio at 1% . What is important in this comparison is how effective ElectroShape is at finding molecules with distinct chemotypes. The way to test that is to run the enrichment study with an active set consisting only of one active from each chemotype: this measures only the cross-chemotype enrichments. Ideally, one would want to run this test for every possible such active subset, and average the results.

The combinatorics precluded the possibility of doing this on all 13 datasets. For instance, the EGFR actives are grouped into forty different chemotypic groups, each group containing up to 183 actives. Thus, instead the active subset was assembled by selecting one member from each chemotype at random, repeated 100, 000 times, and averaging the enrichments obtained. The number 100, 000 was chosen as there were no changes in the first three significant figures of the result when the procedure was repeated.

Ultimately, the averaged results across the thirteen datasets gave an enrichment ratio at 1% of 9.5, a decrease of 14% from 11.1. This is a relatively modest decrease, and demonstrates that ElectroShape provides more than simple chemotype matching.

Adding a fourth moment

Cannon et al. [29] demonstrated that USR could be improved by adding a fourth statistical moment to the descriptors, although it was generally worsened by the addition of a fifth moment. Therefore, we investigated the effect of adding a fourth moment (the fourth root of the fourth central moment) to ElectroShape, USR and CSR, to test if a similar improvement could be obtained. The comparison was done using the enrichment ratios at 1% and the DUD datasets with the various charge sets; the results are presented in Table 3.

As can be seen, though the addition of a fourth moment improves USR and CSR, ElectroShape does not demonstrate any consistent trend. Although it might seem that adding extra information could only improve enrichment,

Table 3 Comparison of enrichment ratios at 1% ($E_{1\%}$) for USR, CSR and ElectroShape (at a ratio of 25 Å per electron charge, see Sect. 4.1), with three or four moments in the descriptor, averaged across the entire DUD release 2 datasets

Method	Charges	$E_{1\%}$ (3 moments)	$E_{1\%}$ (4 moments)	Difference
USR	N.A.	7.4	7.6	0.2
CSR	N.A.	7.7	7.8	0.1
ElectroShape	AM1	10.8	10.8	0.0
	Gasteiger	10.8	11.0	0.2
	AMSOL OPT	12.6	12.5	−0.1
	MMFF94x	13.3	13.1	−0.2
	DUD original	30.2	30.3	0.1

this is not necessarily true. There is a balance between adding extra information and diluting the information already included. Weighting the extra moment equally to the first three dilutes the information in those three. If there is an overlap between the information in the fourth moment and the information already included, this can skew the results and worsen the enrichment. Weighting the fourth moment differently to the others, by some constant factor, is possible, but risks overfitting to the data.

The results presented here and in [29] imply that for spatial USR-type methods, the balance between extra information and dilution lies around the fourth moment, whereas for ElectroShape, it lies between the third and fourth moments.

Conclusion

The four dimensional electrostatic and shape recognition program, ElectroShape, has excellent enrichment performance, and has been shown to be a very useful tool in virtual screening. It combines the computational efficiency of USR with a richer molecular description. Its performance is significantly better than the Chiral Shape Recognition (CSR) method [17], which is itself better than USR [14, 15]—these had average enrichment ratios at 1% of 7.7 and 7.4, respectively, while ElectroShape's average enrichment ratio at 1% is 13.3 with MMFF94x partial charges. This is a near doubling in performance over USR and CSR.

Furthermore, the approach—using supplementary numerical information as an extra coordinate, and applying shape recognition algorithms to it—has many generalisations, and could result in a whole new family of virtual screening methods. For instance, if vector-valued information was attached to each atom, such as electrostatic field information, then the whole molecule could be seen as a collection of points in six dimensional space, and the shape recognition program adjusted accordingly. It is clear from the above that by extending molecular similarity to higher dimensions, there are considerable advantages and opportunities.

References

1. Leach AR, Gillet VJ, Lewis RA, Taylor R (2010) Three-dimensional pharmacophore methods in drug discovery. J Med Chem ASAP
2. Köppen H (2009) Virtual screening: what does it give us?. Curr Opin Drug Discov Dev 12(3):397–407
3. Shoichet BK (2004) Virtual screening of chemical libraries. Nature 432:862–865
4. Kubinyi H (2006) Pharmaceutical research and development
5. NCI Diversity Set II, http://www.dtp.nci.nih.gov/branches/dscb/div2_explanation.html
6. Nicholls A, McGaughey GB, Sheridan RP, Good AC, Warren G, Mathieu M, Muchmore SW, Brown SP, Grant JA, Haigh JA, Nevins N, Jain AN, Kelley B (2010) Molecular shape and medicinal chemistry: a perpective. J Med Chem. doi:10.1021/jm900818s
7. Willett P, Barnard JM, Downs GM (1998) Chemical similarity searching. J Chem Inf Comput Sci 38:983–996
8. Kortagere S, Krasowski MD, Ekins S (2009) The importance of discerning shape in molecular pharmacology. Trend Pharmacol Sci 30:138–147
9. Patterson DE, Cramer RD, Ferguson AM, Clark RD, Weinberger LE (1996) Neighborhood behavior: a useful concept for validation of molecular diversity descriptors. J Med Chem 39:3049–3059
10. Hawkins PCD, Skillman AG, Nicholls A (2007) Comparison of shape-matching and docking as virtual screening tools. J Med Chem 50(1):74–82
11. Grant JA, Gallardo MA, Pickup B (1996) A fast method of molecular shape comparison: a simple application of a gaussian description of molecular shape. J Comp Chem 17(14):1653–1666
12. Rush TS, Grant JA, Mosyak L, Nicholls A (2005) A shape-based 3-D scaffold hopping method and its application to a bacterial protein-protein interaction. J Med Chem 48(5):1489–1495
13. Sheridan RP, McGaughey GB, Cornell WD (2008) Multiple protein structures and multiple ligands: effects on the apparent goodness of virtual screening results. J Comput Aided Mol Des 22(3-4):257–267
14. Ballester PJ, Finn PW, Richards WG (2009) Ultrafast shape recognition: evaluating a new ligand-based virtual screening technology. J Mol Graph Model 27:836–845
15. Ballester PJ, Richards WG (2007) Ultrafast shape recognition to search compound databases for similar molecular shapes. J Comput Chem 28:1711–1723
16. Ballester PJ, Westwood I, Laurieri N, Sim E, Richards WG (2010) Prospective virtual screening with ultrafast shape recognition: the identification of novel inhibitors of arylamine n-acetyltransferases. J R Soc Interface 7(43):335–342
17. Armstrong MS, Morris GM, Finn PW, Sharma R, Richards WG (2009) Molecular similarity including chirality. J Mol Graph Model 28:368–370
18. Huang N, Shoichet BK, Irwin JJ (2006) Benchmarking sets for molecular docking. J Med Chem 49(23):6789–6801
19. Hawkins GD, Giesen DJ, Lynch GC, Chambers CC, Rossi I, Storer JW, Li J, Zhu T, Thompson JD, Winget P, Lynch BJ, Rinaldi D, Liotard DA, Cramer CJ, Truhlar DG (2003) AMSOL-version 7.1. University of Minnesota, Minneapolis
20. Liotard DA, Healy EF, Ruiz JM, Dewar MJS (1989) AMPAC-version 2.1
21. Gasteiger J, Marsili M (1980) Iterative partial equalization of orbital electronegativity—a rapid access to atomic charges. Tetrahedron 36:3219–3288
22. Dewar MJS, Zoebisch EG, Healy F, Stewart JJP (1985) AM1: a new general purpose quantum mechanical molecular model. J Am Chem Soc 107(13):3902–3909 doi:10.1021/ja00299a024
23. Halgren TA (1996) The merck force field. J Comp Chem 17:490–641
24. Chemical Computing Group, Montreal, Canada, MOE 2008.10 http://www.chemcomp.com
25. Armstrong MS, Morris GM, Finn PW, Sharma R, Moretti L, Cooper RI, Richards WG, DUD datasets with new partial charges, http://www.inhibox.com/dud
26. Cheeseright TJ, Mackey MD, Melville JL, Vinter JG (2008) Fieldscreen: virtual screening using molecular fields: application to the DUD data set. J Chem Inf Model 48(11):2108–2117

27. DUD: A Directory of Useful Decoys, http://www.dud.docking.org/r2/ accessed on 31 March 2010
28. Good AC, Oprea TI (2008) Optimization of CAMD techniques 3. Virtual screening enrichment studies: a help or hindrance in tool selection?. J Comput Aided Mol Des 22:169–178
29. Cannon EO, Nigsch F, Mitchell JBO (2008) A novel hybrid ultrafast shape descriptor method for use in virtual screening. Chem Central J 2(3), doi:10.1186/1752-153X-2-3