ELSEVIER

Molecular fields in drug discovery: getting old or reaching maturity?

Simon Cross[1] and Gabriele Cruciani[2]

[1] Molecular Discovery Limited, 215 Marsh Road, Pinner, Middlesex, London HA5 5NE, United Kingdom
[2] Laboratory for Chemometrics and Cheminformatics, Chemistry Department, University of Perugia, Via Elce di sotto 10, I-06123 Perugia, Italy

With GRID first published 23 years ago, and CoMFA 20 years ago, the two most widely known methods that apply molecular fields to drug discovery are now into their third decade. Are molecular-field-based methods still applicable to modern drug discovery? Are they old and outdated? Or are they maturing into their full potential?

Introduction

Drug discovery is a complex process that is costly and takes many years [1]. Typically, once a therapeutic target is identified, and appropriate assays developed to enable the *in vitro* testing of potential drugs, the identification of potential drugs can begin. Millions of compounds are available commercially or in-house, and subsets of these can be assayed to find 'hits' that show activity in the assay, using high-throughput screening (HTS). After confirmation through secondary assays (e.g. dose–response) they are prioritised according to various criteria, including chemical tractability, intellectual property, physicochemical properties and potency. Through an iterative follow-up process, analogues of the best 'leads' are made with the aim of improving potency, reducing off-target effects, obtaining favourable pharmacokinetic and metabolism profiles and avoiding toxicity. Lead compounds that exhibit favourable pharmacodynamic and pharmacokinetic profiles can then be prioritised according to their efficacy *in vivo* and one or more 'drug candidates' chosen for clinical testing. Many years of clinical testing follow before successful candidates are approved for use.

In addition to the experimental *in vitro* and *in vivo* approaches, computer simulation (coined as *in silico*) is now routinely used as a tool to prioritise experiments at each stage of the process [2]. The later that a compound fails in the discovery process, the more costly it has been, hence predicting this failure earlier in the process is highly desirable. Virtual compounds can be filtered according to calculated and predicted physicochemical properties to increase their chances of exhibiting favourable pharmacokinetic properties [3] and chemical tractability. Virtual screening can be performed with many different methods, to prioritise compounds in terms of potency and selectivity so that the more promising compounds are tested [4]. Focused libraries around the most promising hits are then designed; once these analogues have been tested, structure–activity relationships (SAR) may be found *in silico* and statistical methods used to build quantitative models (QSAR), enabling future virtual analogues to be prioritised before synthesis [5]. Structure-based design methods may be used directly on experimentally determined ligand–protein complexes or apo-structures [6]; docking potential ligands into target structures is one method that can be used for virtual screening, and knowledge of co-crystallised ligands can improve this method [7]. In the absence of experimentally determined structures comparative modelling of the target structure may also be possible if suitable homologues are available [8]. Perhaps more simplistically, pure *in silico* visualisation of a ligand–receptor complex can help in understanding the SAR of the system by identifying accessible binding pockets, especially if they are unique to the target and hence provide an opportunity for improving the selectivity profile. Alongside potency and selectivity, pharmacokinetic properties can be optimised using *in silico* quantitative structure–property relationship (QSPR) models [9]. More specifically, methods are available to enable the optimisation of metabolic stability, metabolite prediction, and also to predict toxicity [10]. Whilst *in silico* methods have been developed to support many aspects of the drug discovery process, it is important to stress that they are no substitute for experiment and have limitations; they are a guide to help prioritise the vast number of experiments available and understand the results.

Corresponding author: Cross, S. (Simon@moldiscovery.com)

1359-6446/06/$ - see front matter doi:10.1016/j.drudis.2008.12.006

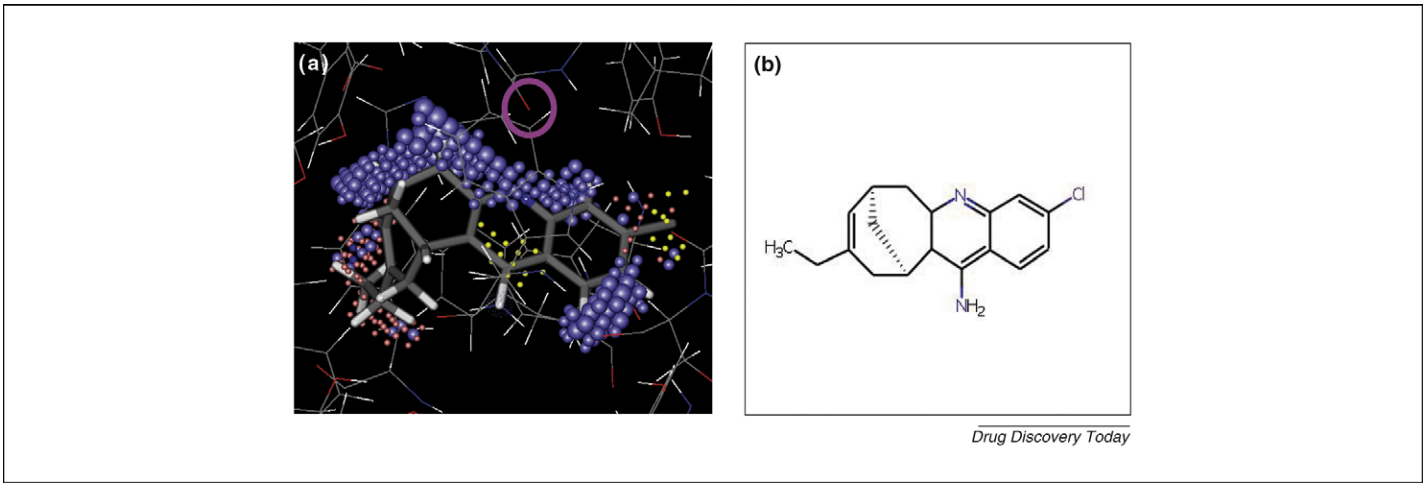

FIGURE 1

(a) Huprine X in complex with acetylcholinesterase. The potentially interacting carbonyl is highlighted with the magenta circle, the nitrogen in the quinoline ring partially obscured by the blue GRID field points representing favourable hydrogen-bond donor interactions (N1 probe). **(b)** 2D structure of Huprine X.

Of course, there are far too many diverse *in silico* approaches to review in one paper; here we focus on methods that describe chemical structures in terms of their molecular fields, with GRID [11] and CoMFA [12] being two of the widest known technologies using this approach. We define a molecular field as a distribution of potentials of physical interaction; including properties such as electrostatic, hydrogen-bonding, hydrophobic and van der Waals interactions. Molecular fields thus describe how a molecule 'appears' to external observers; if two diverse chemical structures are interacting with a target receptor at the same site they will be making a similar set of interactions with the protein, presenting similar molecular fields. Since pharmaceutical discovery is also primarily about finding novel chemical structures, molecular field methods are particularly useful as different structures can exhibit similar fields, thus looking for similar fields can return these different relevant chemotypes.

In the following sections we discuss examples of how these descriptors have been used for a wide range of the other *in silico* applications introduced above.

Virtual compound preparation and filtering

For *in silico* methods, compounds are usually represented as a set of atoms and bonds in a single structural representation. Whilst sounding fairly obvious and sensible, this gives rise to several problems. Have the structures been sketched by different chemists who may represent the same functionality differently? Should tautomers be treated differently? Should I consistently ionise or neutralise my structures, or try to represent them accurately for each individual case depending on the relevant pH? Perhaps confusingly, the answer most probably depends on the computational method being applied, and how the structures were prepared during its parameterisation and validation. Consider one such example shown in Fig. 1 [(a) with Huprine X bound to acetylcholinesterase (PDB: 1e66) and (b) as the 2D representation]. Many computational approaches will represent the quinoline with the nitrogen unprotonated, however there is a carbonyl in close proximity (2.91 Å, within hydrogen bonding distance), suggesting that the nitrogen is protonated when interacting with the receptor. Computational approaches subsequently used on the unprotonated structure may therefore give inaccurate results, depending on their parameterisation, especially structure-based methods such as docking that will score the compound based on interactions with the receptor.

There are several approaches addressing this compound preparation problem. Recently, however, a new method called MoKa has been published which is based on GRID molecular interaction fields (MIFs) [13]. Ionisable centres are defined as fragments in terms of their MIFs, as well as the MIFs of the atoms connected to them through increasing numbers of bond lengths (thus including information on longer range electronic effects), illustrated in Fig. 2. These descriptors are represented in binary fingerprints and were used on a dataset including 25,000 experimental pK_a measurements, along with binary PLS to build a set of quantitative models for 35 ionisation centre types. The model standard error was 0.4 log units, extending to 0.7 log units for novel structures, which compares favourably with 0.1–0.3 log units for the typical experimental error. When compared with ACD/pK_a [14] on a more challenging dataset containing drug-like moieties and not available in the literature, MoKa gave a standard error of 0.9 log units, ACD/pK_a gave a standard error of 1.4 log units. For the example above, MoKa predicts the pK_a of the quinoline nitrogen in Huprine X to be 9.7, hence it would be protonated under physiological conditions. Of course to be useful in a wider context, the method needs to be able to run automatically and be able to process large numbers of structures; currently MoKa is able to process ~1 million structures per hour. Whilst not the first attempt to apply molecular-field-based methods to pK_a prediction, it is the first that does it in a global and high-throughput context. Importantly, where limitations are found applying it to certain challenging chemical series, experimental data can quickly be added to refine the global model. In addition to protonation state, tautomer representation can also be important when using *in silico* methods, and molecular-field-based pK_a prediction has been used to predict the predominant forms in aqueous solution [15].

After structural representations of compounds have been prepared, it has become widespread practice in the pharmaceutical

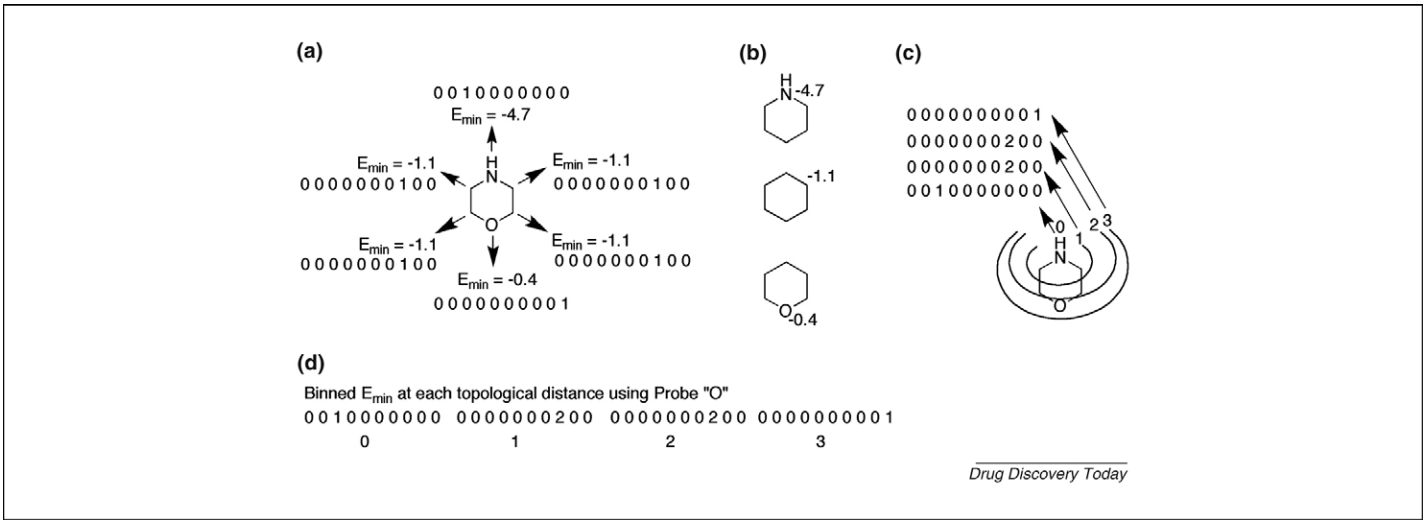

FIGURE 2

Example of the workflow used to generate the descriptors in MoKa: **(a)** binned energies used to describe each atom of morpholine using **(b)** fragments and GRID MIF energy minima (kcal/mol) calculated for the reference atoms using probe 'O'. **(c)** Each bin is summed across the atoms at the same level and concatenated to give the fingerprint descriptors for each probe type shown in **(d)**.

industry to filter them with the aim of removing undesirable functionality, and applying various rules to increase the chance of success further down the pipeline of discovery (i.e. focus efforts on drug-like compounds). The most widely quoted of these, Lipinski's rule-of-5 [3], is aimed at avoiding poor absorption and permeation. A key descriptor is the partition coefficient, log *P*, that describes the lipophilicity of compounds. Given that 95% of drugs may be ionisable, and that log *P* applies to the neutral species of a compound, it may be preferable to use the pH-dependent distribution coefficient, log *D*, which is related to log *P* through pK_a. We have already seen how molecular fields can be applied to estimating pK_a with good accuracy, but what about log *P* and log *D*? Using ~14,000 experimental log *P* measurements, a model was built using the GRID-derived VolSurf [9] descriptors, with a cross-validated r^2 of 0.95, and a standard error of 0.5 log units, close to the experimental error of 0.4 log units. So in two key aspects of modern *in silico* discovery, molecular field approaches can play an important role.

Molecular fields and virtual screening

As discussed above, virtual screening enables experimental screening to focus on compounds that are more likely to be active. It is routinely applied in pharmaceutical discovery and typically encompasses many different algorithms that use receptor information (receptor-based virtual screening) or known active compound information (ligand-based virtual screening). Some methods can employ a combination if the information is known (e.g. using a ligand co-crystallised in a receptor). Molecular field approaches have been used previously in this area with the field-based similarity search (FBSS) approach [16] from Wild *et al.* using molecular electrostatic potentials as descriptors, a genetic algorithm to maximise the alignment between a pair of structures and the Carbo index to evaluate their similarity. More recently, Tervo *et al.* have described another variant on this method (BRUTUS [17]) that uses steric and electrostatic fields to align rigid structures yielding some virtual screening success. The MIMIC [18] algorithm proposed by Mestres *et al.* also uses the Carbo index to compare steric and electrostatic field similarity, and has been applied to virtual screening as part of the SP-DOCK and SG-DOCK programs [19]. In this case, the solutions from the docking algorithm were modified during the docking process, according to the molecular-field-based similarity of the solutions to a co-crystallised ligand. The molecular-field-corrected approaches performed better than the pure receptor-based docking alone. In the remainder of this section we will focus on three specific methods that use molecular fields for virtual screening that have shown retrospective and prospective success and yet are very different in their approaches.

Following on from one of the most well known molecular field approaches, the topomer method [20–25] brings the steric field from CoMFA [12], along with some pharmacophore description, together with the popular recent paradigm of working with molecular fragments. Invariant 3D representations of these fragments are generated by applying a set of deterministic rules; hence equivalent fragments from different structures are aligned according to these same rules and their fields comparable (see Fig. 3). What is striking about this method is that the whole molecule, when rebuilt from the fragments, may be completely implausible and contain steric clashes or unrealistic internal geometries [25]; the goal is consistency over realism. The authors have shown that the descriptors exhibit better neighbourhood behaviour than other metrics including 2D fingerprint similarity. Prospective performance was examined by 352 queries from World Drug Alert records; 40% of these yielded one or more virtual hits when using a distance cutoff, and represented over 50 different activity classes. Owing to pragmatic considerations (cost of assay, material available and query activity below threshold) only 308 compounds were tested in 13 different target assays. Randomly selected compounds chosen from the test library showed an average hit rate of 2% (90% inhibition at 10 μM) and for the topomer selected compounds 21%. Additionally, for 11 of the 13 targets, the solutions were different structurally from the queries (the average Tanimoto 2D fingerprint similarity was 0.36) and hence good examples of

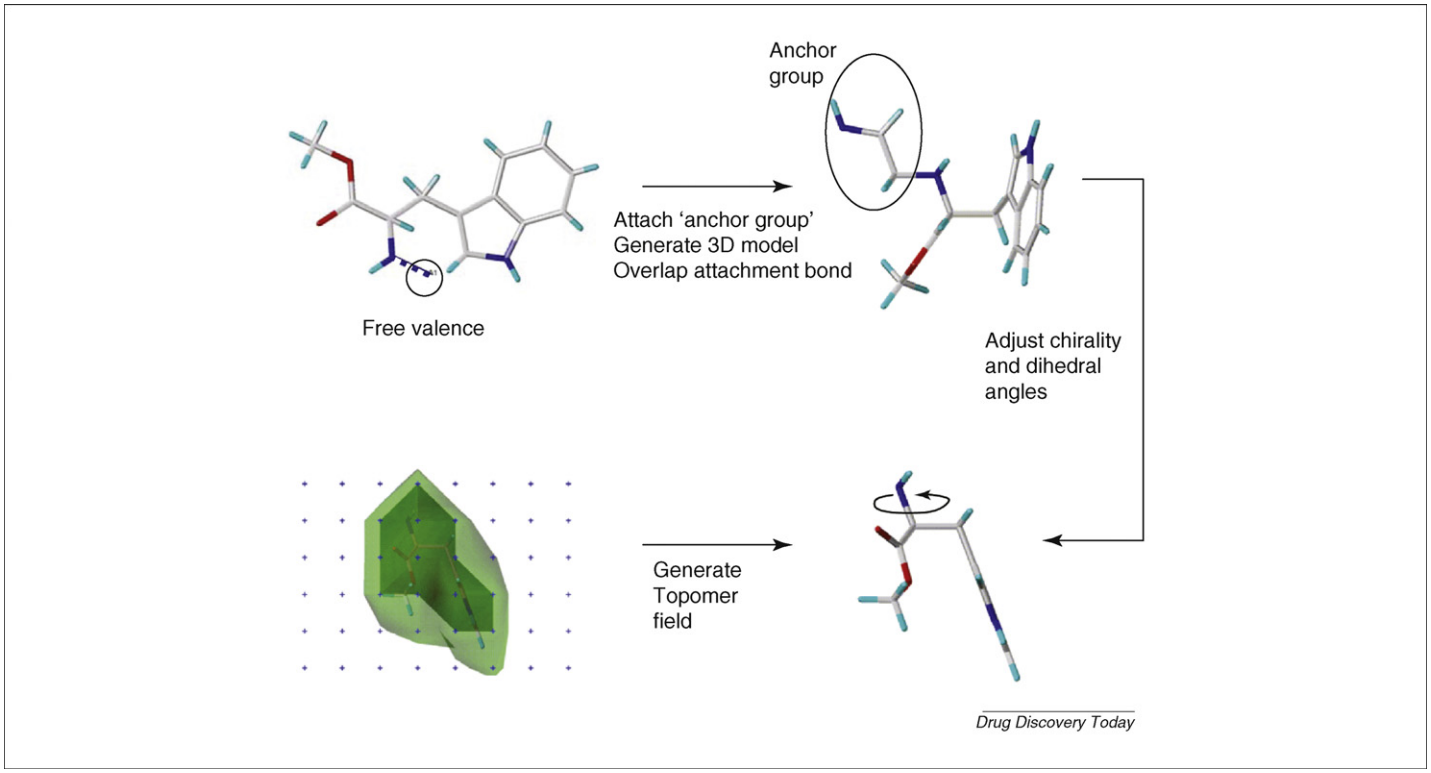

FIGURE 3

Steps in the generation of a topomer from a 2D chemical fragment. Figure kindly provided, with permission, by Cramer *et al.* [21].

how molecular field approaches can provide more chemical diversity than other methods. That it is based on molecular fragments lends it to the searching of larger chemical libraries consisting of virtual but synthetically accessible compounds; the authors report that 10^{20} different structures are available to the method [26], an impressive number considering that virtual screening is usually performed using, at most, millions of available compounds.

At the other end of physicochemical realism, Vinter *et al.* use non-atom centred charges in *pi* systems in the XED forcefield [27,28]; electrostatic, shape and hydrophobic fields derived from this better explain crystallographic data than using the simpler approximations offered by atom centred charges [29]. The fields are determined using a probe atom with a charge of +1, 0 or −1, field points are then placed at the extrema of the interactions. The exception is the hydrophobic field, where field points are positioned at the centre of hydrophobic groups such as phenyl, halogens and alkyl (see Fig. 4). These field points are then used to describe molecules and compare their similarity, one application of which is virtual screening. Once a template field has been generated (co-crytallised ligand, docked ligand and alignment hypothesis), it can be used to screen a database containing test structures with multiple conformations and their associated fields. In terms of retrospective validation, the FieldScreen method has recently been applied to the DUD dataset [30]. Out of the 40 target datasets, 13 were chosen that included at least 15 clusters of active chemotypes. Using awROC enrichment at 1%, for 9 of the 13 targets it performed better than DOCK [31]. The method has also been used both in a scaffold-hopping and virtual screening approach to find CCK2 antagonists [32,33]. In this case, a known lead with an undesirable core was used as a template, and the XED field approach used to find novel scaffolds that mimicked the field pattern around the undesirable core. Two such scaffolds were reported, and the structures synthesised and found to be active. Subsequently, a 600,000 commercial compound database was searched using two individual leads, modelled in their bioactive conformation, as templates. Of the top 500 compounds from each search, 88 were purchased and tested; 27 had greater than 20% binding at 10 μM (30% hit rate), and 4 had greater than 20% binding at 1 μM. Importantly, the reported structures were structurally distinct from the initial templates, and other structures were unreported for commercial reasons.

The GRID forcefield [11] was specifically designed to describe molecular interactions and, in particular, with biomolecules such as drug targets. It is calibrated empirically and, rather than computing the more abstract electrostatic, steric and hydrophobic fields, it describes these effects through interactions of the target under investigation with over 60 chemical probes. Some have highlighted limitations of GRID-based methods for virtual screening: that either too many points are produced for fast processing or accuracy has to be sacrificed by using fewer points on a lower resolution grid or that small rotations/translations within the grid give inconsistencies in the results (gauge variance). Despite these objections, GRID-based methods have been successfully applied, as the examples below demonstrate. One particular application of the GRID forcefield has been published recently by Baroni *et al.* as the FLAP algorithm [34]. With FLAP, several GRID probes are used to describe their MIFs with the target, which can be a receptor or a small molecule. These MIFs are then condensed into discrete pharmacophoric points representing favourable and unfavourable interactions using a weighted energy-based and space coverage

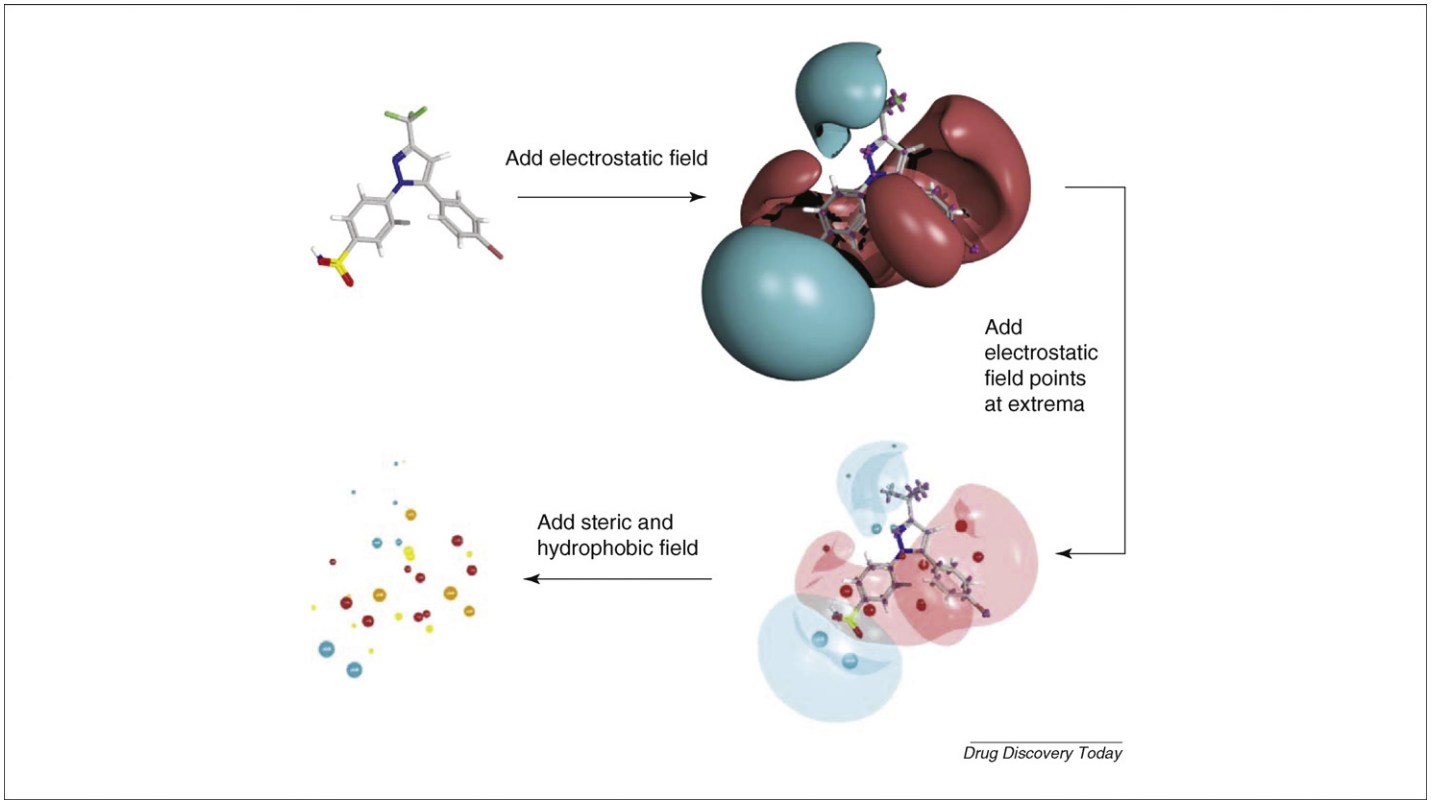

FIGURE 4

Steps in the conversion of a 3D structure to a XED field representation. Red denotes positive electrostatic field points, blue denotes negative electrostatic points, yellow denotes van der Waals attractive points and orange denotes hydrophobic points. Figure kindly provided, with permission, by Cheeseright *et al.* [29].

function, and these can also be manually edited by the user to focus on specific interactions. All quadruplets are generated; potential ligands are then tested to see if their atoms can match up with these quadruplets after conformational searching and receptor site shape filtering (see Fig. 5). The energy of interaction of these quadruplet matches and the number of these favourable events are used to rank large numbers of compounds for virtual screening. The MIF sampling overcomes the limitation of speed versus accuracy, hence grids of high resolution can be used. At this resolution, the gauge variance effect is minimal and bypassed by the tolerances that are allowed when matching the quadruplets. Since the receptor is in reality flexible and conformational sampling is also an approximation, such tolerances are needed to allow for this error. Of course the real question is whether the method performs well. In a receptor-based screening approach retrospective screening was carried out on Factor Xa and thymidine kinase using known actives and decoys chosen from MDDR. These initial results demonstrated that the FLAP approach was comparable with other current docking-based approaches. In a ligand-based screening approach, FLAP was applied to a project at Pfizer and compared favourably with a well-known pharmacophore method. Another comparison was performed on an estrogen receptor dataset along with MDDR decoys, comparing FLAP with other pharmacophore and 2D fingerprint methods. Again the method performed well in comparison, and all of the known actives were returned within only 2% of the dataset. A more rigorous virtual screening

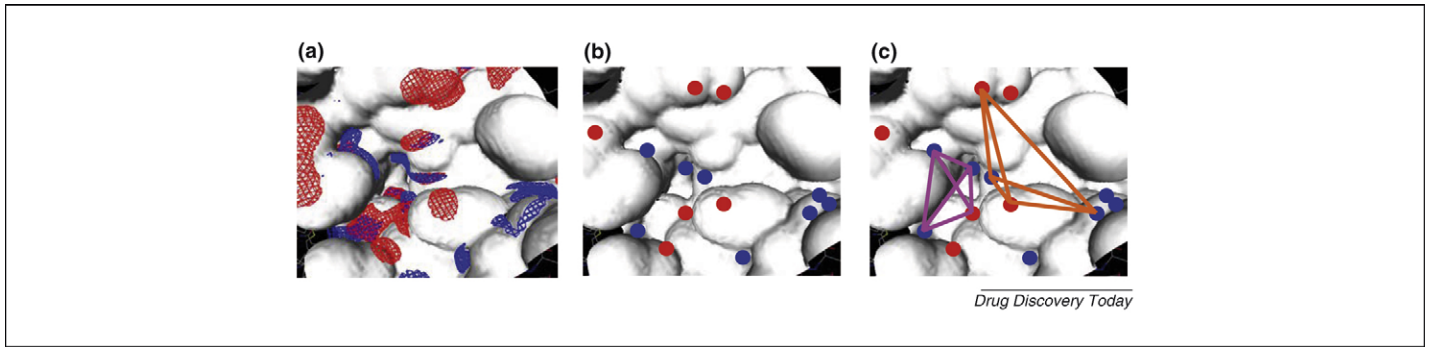

FIGURE 5

FLAP method. **(a)** Molecular interaction fields (MIFs) are calculated in the active site of a protein structure. **(b)** The MIFs are condensed into a few target-based pharmacophoric points. **(c)** All possible arrangements of four pharmacophoric points are generated.

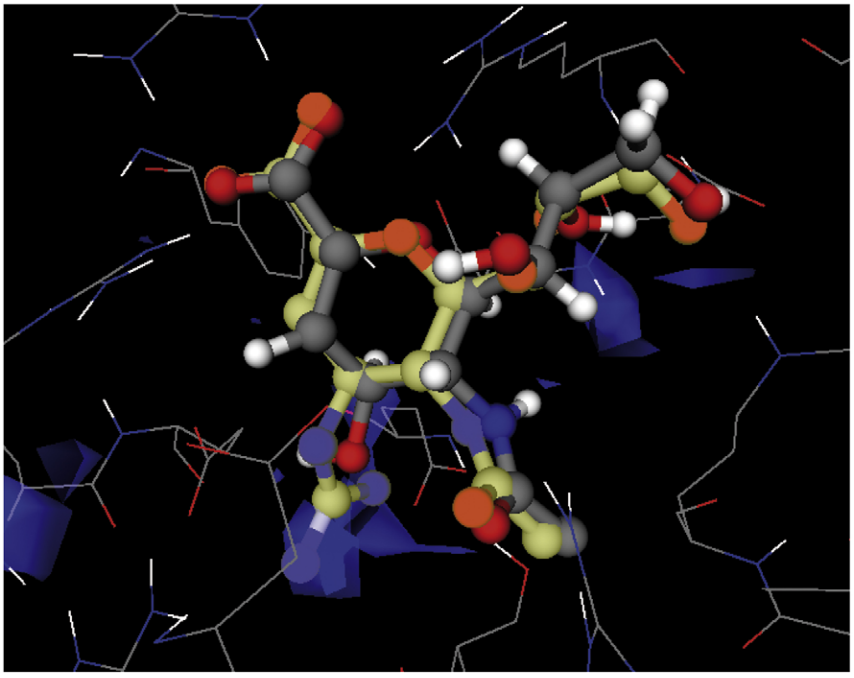

FIGURE 6

Neuraminidase in complex with DANA (grey carbons) and Zanamivir (atoms highlighted). The blue contours indicate the favourable interactions of the protein with a guanidinium probe using GRID.

validation study is currently underway using the Directory of Useful Decoys (DUD) [30] with 40 targets and the DUD-self decoys that were chosen to have similar properties to the known actives and hence be a much tougher test. Preliminary results are encouraging, with average enrichment factors at 1% of 24.9 (maximum of ~38.5) and average ROC AUC values across all targets as high as 0.94 (maximum of 1.0) (pers. commun.). In a recent prospective study, ligand-based FLAP has recently been applied in a multi-disciplinary screening cascade to find calcium channel antagonists from a starting point of 350,000 purchasable compounds. Twenty compounds were finally selected and tested, leading to the discovery of three novel active chemotypes [35]. Recently, an alternative method based on GRID MIFs has been described for scaffold-hopping [36]. Using anchor-GRIND descriptors (and thus also bypassing the GRID limitations described above), scaffolds are described in terms of their GRID fields using a ligand-based or receptor-based perspective. The method was validated by demonstrating recovery of thrombin, HIV-protease and neuraminidase scaffolds from a decoy database of known combinatorial library scaffolds. These examples demonstrate the value of using molecular-field-based methods to virtual screening; the methods vary in their implementation, but all have shown retrospective, prospective and practical success in identifying promising hits.

Molecular fields and SAR

Once active structures are found against the target of interest, project work typically focuses on improving potency, selectivity and various pharmacokinetic properties (discussed in the following section). One of the earliest and most significant examples of the benefits of rational or receptor-based design was demonstrated for the target neuraminidase [6], using the GRID method introduced in the previous section. The crystal structure of neuraminidase in complex with an unsaturated sialic acid analogue (DANA, $K_i = 1$ μM) was solved, and GRID used to analyse the active site to predict favourable interactions. A strong interaction was seen leading to the replacement of the 4-hydroxyl by the 4-guanidino (see Fig. 6) which was 5000 times more potent ($K_i = 0.2$ nM) and was subsequently developed as Zanamivir.

In addition to this simple but powerful approach to optimising interactions with the target receptor, selectivity studies have been carried out, highlighting structural differences between the target and other similar receptors. This generally involves superimposition of the binding sites, GRID MIF calculation for each target, statistical analysis using PCA or consensus PCA, then interpretation of the results. Such selectivity analyses have been applied to diverse systems such as DNA minor groove binding [37], DHFR [38], COX-1/COX-2 [39], penicillin acylase [40,41], serine proteases [42], CYP450 [43], protein kinases [44], matrix metalloproteases [45], nitric oxide synthases [46], PPARs [47], bile acid transportation [48] and Eph kinases [49] (reviewed in [50]).

For ligand-based SAR modelling using molecular fields, perhaps the most well known method is comparative molecular field analysis (CoMFA) [12], which compares the steric and electrostatic molecular fields between a series of aligned molecules using PLS statistical analysis. The key advantage in using molecular fields is that their contributions to the model can be directly visualised in 3D, highlighting where changes can be made to improve potency, and the power of the approach is indicated by the hundreds of

references that can be found applying the approach successfully to drug design, as well as many extensions or modifications of the original method (e.g. CoMSIA [51], HINT [52], GRID/GOLPE [53], HASL [54], COMPASS [55] and AFMoC [56]). Perhaps the single most important drawback for CoMFA analyses is the sensitivity to the alignment of the input structures, and usually a structural scaffold is required to make this consistent, which also makes CoMFA most applicable within congeneric series.

To avoid the alignment problem, Pastor *et al.* developed the Grid independent (GRIND) descriptors approach [57], whereby GRID MIFs for each molecule are filtered by energy and MIF node pair distance. The GRIND descriptors represent the product of the interaction energy for each pair of nodes, binned by distance, with only the maximum products being retained. After statistical analysis such as PLS, the QSAR model is obtained and the strongest contributions can be traced back to the node pairs (analogous to pairs of pharmacophore points) which can then give insight into improving activity. Recently a modification that increases the specificity of the descriptors has also been described [58], and also improvements to the MIF point sampling (pers. commun.). It should be noted that in such an approach finding appropriate molecular conformations is still a challenging task. Additionally, another GRID-based method has been developed that compresses the 3D GRID maps into a few quantitative 2D descriptors; typically more relevant for ADME modelling, it has also been used successfully to model the SAR of anti-HIV quinolones [59].

The topomer descriptors discussed above have also been used in the CoMFA context, solving the alignment and conformation problems; the topomers are generated by deterministic rules that consistently describe fragments in 3D and aligned according to predetermined rules. This, however, does require the compounds to share some kind of common core or 'equivalent' acyclic bond where the structures are split into fragments for comparison. Once the topomers have been generated (and with them the alignments), the CoMFA statistical analysis is performed with a few minor modifications to 'standard CoMFA' [60]. Retrospective validation was carried out using 15 datasets chosen from the literature. The topomeric CoMFA models showed an average q^2 of 0.52 (SDEP = 0.69) compared with the literature average of 0.64 (SDEP = 0.55), hence they were generally weaker than the models built by more careful alignment. Even so, it is significant to have such models after just a few minutes work, especially given that the error of prediction is generally ~0.75 kcal/mol, when compared with other (receptor-based) methods where the error is at best 2–5 kcal/mol, and that the topomeric groups under study can be directly used for virtual screening to find more optimal fragments that increase potency.

The approach from Vinter *et al.* to solving the conformational problem is to compare their XED-derived field minima points between pairs of active structures, in each case considering several generated conformations. Cross-correlating these pairs, or duos, from several active molecules yields the bioactive conformation

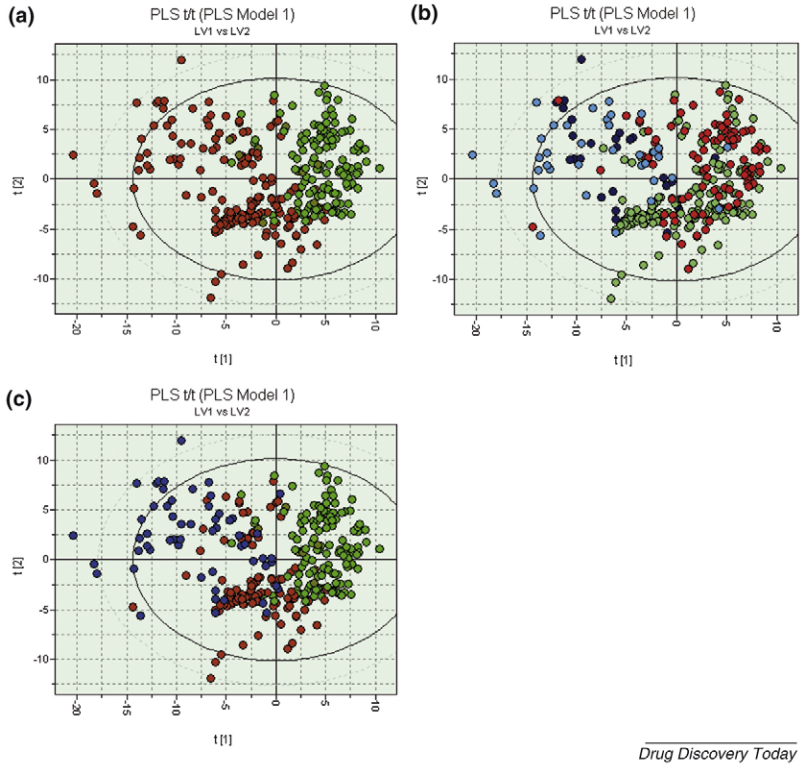

FIGURE 7

(a) VolSurf PLS scores plot showing the separation of compounds that cross (green circles) and do not (red circles) the blood–brain barrier. **(b)** The same compounds are shown and coloured according to whether they are anionic (pale blue), zwitterionic (blue), cationic (red) and neutral (pale green). **(c)** Follows the same colour scheme as (a), except for the compounds with zero available uncharged species at pH 7.4. It can be seen clearly that charge *per se* is not the discriminating factor, but the availability of some neutral fraction that is able to cross the membrane.

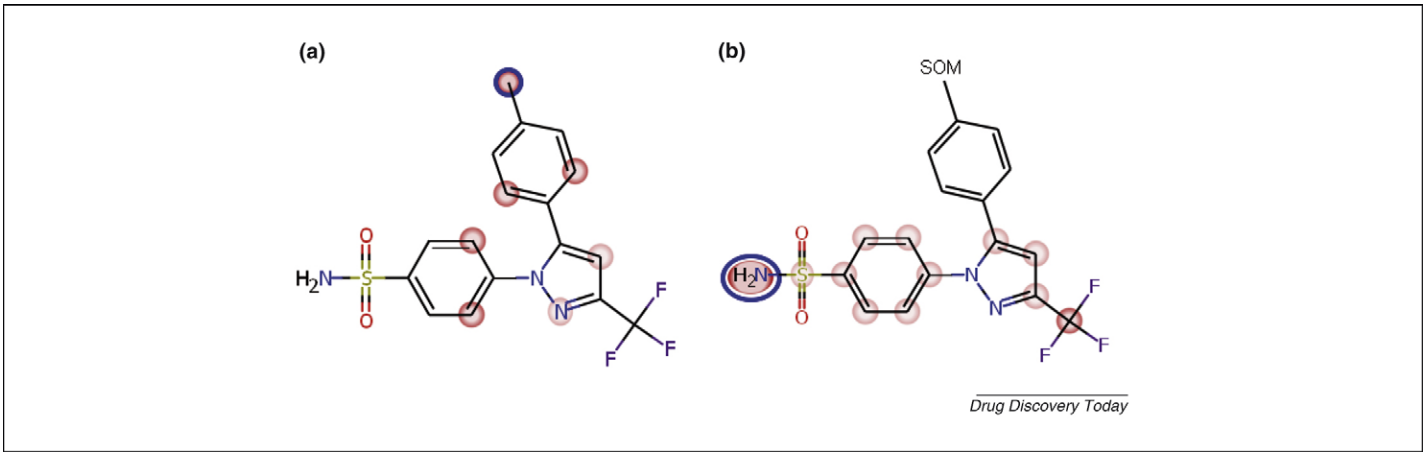

FIGURE 8

MetaSite predictions of the most probable site of metabolism by CYP2C9: **(a)** primary site highlighted with a blue circle and the regions contributing most strongly to the primary benzylic methyl position: **(b)** strongest contributor highlighted with a blue ellipse.

hypothesis. Aligning other active structures to this field template provides a combined field and volume score that has been shown to correlate with activity and exhibit reasonable predictivity [61]. This section has focused on the optimisation of lead compounds with particular respect to potency and selectivity; it is the area most synonymous with molecular field approaches, and the large number of examples highlighted above demonstrate the wide ranging impact they have made.

Molecular fields and ADMET

This last section looks at the complex area of pharmacokinetics, which covers properties such as solubility, permeation, metabolism and the absence of toxicity, which are essential factors in the success of a candidate compound. Molecular field approaches have started to impact this area significantly, enabling *in silico* modelling to help focus the experimental efforts. The GRID-based VolSurf approach introduced in the previous section compresses the 3D MIFs characterising interactions with the water, hydrophobic and hydrogen-bond acceptor probes (OH2, DRY and O) into a smaller number of descriptors particularly relevant for ADME. The descriptors refer to the size and shape, the hydrophilic and hydrophobic interaction regions and the balance between them. These have been described elsewhere in greater detail [9], however recent additions include properties such as the proportion of available uncharged species available at physiological pH, and an automatic search for intramolecular hydrogen bonding and hydrophobic collapse that can affect the conformation of the structures and the subsequent descriptor calculations (pers. commun.; see also Fig. 7: http://www.moldiscovery.com/docs/vsplus/). In the 'Virtual compound preparation and filtering' section we discussed the octanol/water partition coefficient model calculated from VolSurf descriptors; other key ADME-related models that have been built successfully using the same approach include volume of distribution (related to drug half-life), metabolic stability (related to bioavailability, clearance and hence toxicity), passive absorption through PAMPA, Caco-2, MDCK and BBMEC cells, blood–brain barrier permeation [62], oral absorption, unspecific protein binding and solubility [63].

Additionally, GRIND descriptors have been used to model P-glycoprotein (PGP) efflux, CYP 3A4 inhibition, and also HERG inhibition [64], and the GRID/GOLPE and CoMFA approaches have been used to model the PepT1 transport system [65,66], and also the area of CYP inhibition [64]. Whilst all of these models will vary in 'predictivity' and applicability, it seems apparent that molecular field approaches are proving to be valuable. In addition to these QSAR/QSPR approaches, we have also recently described a new approach to predicting the site of metabolism for lead compounds [67], which is one of the most time and material consuming experimental tasks in the ADME field. The MetaSite approach predicts three key aspects that combine to determine the site of metabolism across a range of cytochromes. Firstly, the spatial recognition component uses GRIND descriptors to predict the most probable positions in close proximity to the catalytic heme moiety; given the size, shape and flexibility of the various isoforms, it is important that the GRID method can automatically deal with protein flexibility. Secondly, the chemical reactivity at each position in the molecule is estimated using pre-calculated *ab initio* data on a diverse range of chemical fragments. Finally, the propensity for different metabolic reaction mechanisms by the different isoforms is included. The method predicts the most probable site of metabolism, the structural regions that impact on the relevant spatial orientation within the site and the most probable metabolites that are produced. A potential lead optimisation example is demonstrated by Ahlström *et al.* using the COX-2 inhibitor celecoxib, which is rapidly metabolised by CYP2C9 [67]. The MetaSite algorithm correctly predicts that the benzylic methyl is the most probable site of metabolism (Fig. 8).

From the medicinal chemist's perspective, the probably next step is to block the site with the trifluoromethyl derivative. This improves the stability of the compound to the point where it potentially inhibits the enzyme (leading to toxicity problems). The prediction of which regions of the structure affect its orientation within the cavity provide an alternative modification point. With celecoxib, the sulfonamide moiety was the strongest contributor according to MetaSite, and modification in this position reduced metabolism significantly, yet not enough for it to inhibit the enzyme.

Concluding remarks

With drug discovery becoming increasingly costly and productivity waning, it is clear that *in silico* approaches can benefit the process, guiding experiments to achieve similar results with less effort. Molecular-field-based methods were at the forefront of the computer-aided design discipline, and today are more valuable than ever. As an aid to computer-aided drug design itself, we have seen their impact on compound preparation and filtering. More directly impacting discovery, several field-based methods have been developed for hit-finding by virtual screening; three of these (topomers, FieldScreen and FLAP) in particular are actively used in the pharmaceutical industry. For finding SAR, the two forefathers of field-based methods, GRID and CoMFA, are still actively contributing to the optimisation of potency and selectivity, and both have led to improvements via offspring methods (GRIND and topomer CoMFA). In the area of predictive pharmacokinetics, GRID-derived methods (VolSurf, GRIND and MetaSite) are making a significant impact in one of the most challenging areas for drug discovery. It is clear that molecular field approaches have matured into many diverse applications that significantly contribute to drug discovery – so what does the future hold? Of course, like other methods, the majority of improvements will be incremental, for example increasing throughput via improving algorithm efficiencies and parallelisation, or extracting field information differently so as to increase focus on the relevant structural regions. Fragment-based field approaches are now available using the CoMFA, GRID and XED methods, helping to automate lead optimisation. The ADMET area still offers the most challenges, and it is in this area that we are likely to see the next significant advances. Given the success in Phase I cytochrome-mediated metabolism, molecular field approaches should allow the prediction of Phase II metabolism products. Prediction of off-target effects should be possible using receptor-based and ligand-based approaches providing warnings for potential leads. It may also be possible to describe targets related to toxicity (such as hERG and the nuclear receptors including PPARγ), highlighting the moieties in potential leads that are likely to contribute to their toxicity. In all of the molecular-field-based applications described above, with the exception of MetaSite, the fields are calculated on static ligands or receptors, meaning the bioactive or at least consistently generated molecular conformations are required for successful results. GRID allows the generation of flexible fields where the ligand or receptor undergoes conformational change in response to the chemical probes, but as yet they have not been exploited to their full potential across the virtual discovery process. For the cytochromes in MetaSite, using these flexible fields yielded prediction success of ~85% compared to ~65% using static fields [67]; applying field flexibility to other areas could have a dramatic impact.

References

1 Kapetanovic, I.M. (2008) Computer-aided drug discovery and development (CADDD): in silico-chemico-biological approach. *Chem. Biol. Int.* 171, 165–176
2 Clark, D.E. (2006) What has computer-aided molecular design ever done for drug discovery? *Expert Opin. Drug Discov.* 1, 103–110
3 Lipinski, C.A. *et al.* (1997) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.* 23, 3
4 Clark, D.E. (2008) What has virtual screening ever done for drug discovery? *Expert Opin. Drug Discov.* 3, 841–851
5 Lewis, R.A. (2005) A general method for exploiting QSAR models in lead optimization. *J. Med. Chem.* 48, 1638–1648
6 Von Itzstein, M. *et al.* (1993) Rational design of potent sialidase-based inhibitors of influenza virus replication. *Nature* 363, 418–423
7 Cross, S.S.J. (2005) Improved FlexX docking using FlexS-determined base fragment placement. *J. Chem. Inf. Model.* 45, 993–1001
8 Evers, A. *et al.* (2004) Successful virtual screening for a submicromolar antagonist of the neurokinin-1 receptor based on a ligand-supported homology model. *J. Med. Chem.* 47, 5381–5392
9 Cruciani, G. *et al.* (2000) VolSurf: a new tool for the pharmacokinetic optimization of lead compounds. *Eur. J. Pharm. Sci.* 11, S29–S39
10 Cruciani, G. *et al.* (2005) MetaSite: understanding metabolism in human cytochromes from the perspective of the chemist. *J. Med. Chem.* 48, 6970–6979
11 Goodford, P.J. (1985) A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J. Med. Chem.* 28, 849–857
12 Cramer, R.D. *et al.* (1988) Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. *J. Am. Chem. Soc.* 110, 5959–5967
13 Milletti, F. *et al.* (2007) New and original pK_a prediction method using grid molecular interaction fields. *J. Chem. Inf. Model.* 47, 2172–2181
14 *ACDpKa, V. 8. 03*. Advanced Chemistry Development
15 Milletti, F. *et al.* Tautomer enumeration and stability prediction for virtual screening on large chemical databases. *J. Chem. Inf. Model.* Jan 6, 2009, doi:10.1021/ci800340j.
16 Wild, D.J. *et al.* (1996) Similarity searching in files of three-dimensional chemical structures. Alignment of molecular electrostatic potential fields with a genetic algorithm. *J. Chem. Inf. Comput. Sci.* 36, 159–167
17 Tervo, A.J. *et al.* (2005) Optimization of a grid-based similarity function for rigid-body molecular superposition. 1. Alignment and virtual screening applications. *J. Med. Chem.* 48, 4076–4086
18 Mestres, J. *et al.* (1996) MIMIC: a molecular-field matching program, exploiting applicability of molecular similarity approaches. *J. Comput. Chem.* 18, 934–954
19 Fradera, X. *et al.* (2000) Similarity-driven flexible ligand docking. *Proteins: Struct. Funct. Genet.* 40, 623–636
20 Cramer, R.D. *et al.* (1996) Bioisosterism as a molecular diversity descriptor: steric fields of single topomeric conformers. *J. Med. Chem.* 39, 3060–3069
21 Cramer, R.D. *et al.* (1999) Prospective identification of biologically active structures by topomer shape similarity searching. *J. Med. Chem.* 42, 3919–3933
22 Andrews, K.M. *et al.* (2000) Toward general methods of targeted library design: topomer shape similarity searching with diverse structures as queries. *J. Med. Chem.* 43, 1723–1740
23 Cramer, R.D. *et al.* (2002) dbtop: topomer similarity searching of conventional structure databases. *J. Mol. Graph. Model.* 20, 447–462
24 Cramer, R.D. *et al.* (2004) ''Lead hopping''. Validation of topomer similarity as a superior predictor of similar biological activities. *J. Med. Chem.* 47, 6777–6791
25 Jilek, R.J. *et al.* (2004) Topomers: a validated protocol for their self-consistent generation. *J. Chem. Inf. Comput. Sci.* 44, 1221–1227
26 Cramer, R.D. *et al.* (2007) AllChem: generating and searching 10^{20} synthetically accessible structures. *J. Comput. Aid. Mol. Des.* 21, 341–350
27 Vinter, J.G. (1994) Extended electron distributions applied to the molecular mechanics of intermolecular interactions. *J. Comput. Aid. Mol. Des.* 8, 653–668
28 Vinter, J.G. (1996) Extended electron distributions applied to the molecular mechanics of intermolecular interactions. II. Organic complexes. *J. Comput. Aid. Mol. Des.* 10, 417–426
29 Cheeseright, T. *et al.* (2006) Molecular field extrema as descriptors of biological activity: definition and validation. *J. Chem. Inf. Model.* 46, 665–676
30 Huang, N. *et al.* (2006) Benchmarking sets for molecular docking. *J. Med. Chem.* 49, 6789–6801
31 Cheeseright, T. *et al.* (2008) FieldScreen: virtual screening using molecular fields. Application to the DUD data set. *J. Chem. Inf. Model.* 48, 2108–2117
32 Low, C.M.R. *et al.* (2005) Scaffold hopping with molecular field points: identification of a cholecystokinin-2 (CCK2) receptor pharmacophore and its use in the design of a prototypical series of pyrrole- and imidazole-based CCK2 antagonists. *J. Med. Chem.* 48, 6790–6802
33 Cheeseright, T. *et al.* (2007) Molecular field technology applied to virtual screening and finding the bioactive conformation. *Expert Opin. Drug Discov.* 2, 131–144

34 Baroni, M. *et al.* (2007) A common reference framework for analyzing/comparing proteins and ligands. Fingerprints for ligands and proteins (FLAP): theory and application. *J. Chem. Inf. Model.* 47, 279–294

35 Carosati, E. *et al.* (2006) Calcium channel antagonists discovered by a multidisciplinary approach. *J. Med. Chem.* 49, 5206–5216

36 Bergmann, R. *et al.* (2007) SHOP: scaffold HOPping by GRID-based similarity searches. *J. Med. Chem.* 50, 2708–2717

37 Cruciani, G. *et al.* (1994) A search for specificity in DNA–drug interactions. *J. Mol. Graph.* 12, 116–129

38 Pastor, M. *et al.* (1995) A novel strategy for improving ligand selectivity in receptor-based drug design. *J. Med. Chem.* 38, 4637–4647

39 Filipponi, E. *et al.* (2000) Chemometric rationalization of the structural and physicochemical basis for selective cyclooxygenase-2 inhibition: toward more specific ligands. *J. Comput. Aid. Mol. Des.* 14, 277–291

40 Braiuca, P. *et al.* (2003) A homology model of penicillin acylase from *Alcaligenes faecalis* and in silico evaluation of its selectivity. *ChemBioChem* 4, 615–622

41 Braiuca, P. *et al.* (2004) An innovative application of the ''flexible'' GRID/PCA computational method: study of differences in selectivity between PGAs from *Escherichia coli* and a *Providentia rettgeri* mutant. *Biotechnol. Prog.* 20, 1025–1031

42 Kastenholz, M.A. *et al.* (2000) GRID/CPCA: a new computational tool to design selective ligands. *J. Med. Chem.* 43, 3033–3044

43 Ridderström, M. *et al.* (2001) Analysis of selective regions in the active sites of human cytochromes P450, 2C8, 2C9, 2C18, and 2C19 homology models using GRID/CPCA. *J. Med. Chem.* 44, 4072–4081

44 Naumann, T. *et al.* (2002) Structural classification of protein kinases using 3D molecular interaction field analysis of their ligand binding sites: target family landscapes. *J. Med. Chem.* 45, 2366–2378

45 Terp, G.E. *et al.* (2002) Structural differences of matrix metalloproteinases with potential implications for inhibitor selectivity examined by the GRID/CPCA approach. *J. Med. Chem.* 45, 2675–2684

46 Ji, H. *et al.* (2003) Computer modeling of selective regions in the active site of nitric oxide synthases: implication for the design of isoform-selective inhibitors. *J. Med. Chem.* 46, 5700–5711

47 Pirard, B. (2003) Peroxisome proliferator-activated receptors target family landscape: a chemometrical approach to ligand selectivity based on protein binding site analysis. *J. Comput. Aid. Mol. Des.* 17, 785–796

48 Kurz, M. *et al.* (2003) Insights into the bile acid transportation system: The human ileal lipid-binding protein–cholyltaurine complex and its comparison with homologous structures. *Proteins: Struct. Funct. Genet.* 50, 312–328

49 Myshkin, E. *et al.* (2003) Chemometrical classification of ephrin ligands and Eph kinases using GRID/CPCA approach. *J. Chem. Inf. Comput. Sci.* 43, 1004–1010

50 Fox, T. (2006) Protein selectivity studies using GRID-MIFs. In *Molecular Interaction Fields* (Cruciani, G., ed.), pp. 45–82, Wiley

51 Klebe, G. *et al.* (1994) Molecular similarity indices in a comparative analysis (CoMSIA) of drug molecules to correlate and predict their biological activity. *J. Med. Chem.* 37, 4130–4146

52 Kellogg, G.E. *et al.* (1991) HINT: a new method of empirical hydrophobic field calculation for CoMFA. *J. Comput. Aid. Mol. Des.* 5, 545–552

53 Goodford, P. (1996) Multivariate characterization of molecules for QSAR analysis. *J. Chemom.* 10, 107–117

54 Doweyko, A.M. (1991) The hypothetical active site lattice – in vitro and in vivo explorations using a three-dimensional QSAR technique. *J. Math. Chem.* 7, 273–285

55 Jain, A.N. *et al.* (1994) Compass: predicting biological activities from molecular surface properties. Performance comparisons on a steroid benchmark. *J. Med. Chem.* 37, 2315–2327

56 Gohlke, H. *et al.* (2002) DrugScore meets CoMFA: adaptation of fields for molecular comparison (AFMoC) or how to tailor knowledge-based pair-potentials to a particular protein. *J. Med. Chem.* 45, 4153–4170

57 Pastor, M. *et al.* (2000) GRid-INdependent Descriptors (GRIND): a novel class of alignment-independent three-dimensional molecular descriptors. *J. Med. Chem.* 43, 3233–3243

58 Fontaine, F. *et al.* (2005) Anchor-GRIND: filling the gap between standard 3D QSAR and the GRid-Independent descriptors. *J. Med. Chem.* 48, 2687–2694

59 Filipponi, E. *et al.* (2001) QSAR study and VolSurf characterization of anti-HIV quinolone library. *J. Comput. Aid. Mol. Des.* 15, 203–217

60 Cramer, R.D. (2003) Topomer CoMFA: a design methodology for rapid lead optimization. *J. Med. Chem.* 46, 374–388

61 Low, C.M.R. *et al.* (2008) Rationalizing the activities of diverse cholecystokinin 2 receptor antagonists using molecular field points. *J. Med. Chem.* 51, 565–573

62 Crivori, P. *et al.* (2000) Predicting blood–brain barrier permeation from three-dimensional molecular structure. *J. Med. Chem.* 43, 2204–2216

63 Mannnhold, R. *et al.* (2006) Use of MIF-based VolSurf descriptors in physicochemical and pharmacokinetic studies. In *Molecular Interaction Fields* (Cruciani, G., ed.), pp. 173–196, Wiley

64 Cianchetta, G. *et al.* (2006) Molecular interaction fields in ADME and safety. In *Molecular Interaction Fields* (Cruciani, G., ed.), pp. 45–82, Wiley

65 Zamora, I. *et al.* (2001) Prediction of oral drug permeability. In *Rational Approaches to Drug Design* (Holdje, H.D. and Sippl, W., eds), pp. 271–280, Prous Science

66 Gebauer, S. *et al.* (2003) Three-dimensional quantitative structure-activity relationship analyses of peptide substrates of the mammalian H+/peptide cotransporter PEPT1. *J. Med. Chem.* 46, 5725–5734

67 Ahlström, M.M. *et al.* (2007) CYP2C9 structure–metabolism relationships: optimizing the metabolic stability of COX-2 inhibitors. *J. Med. Chem.* 50, 4444–4452