Molecular interaction fields in drug discovery: Recent advances and future perspectives

ARTICLE *in* WILEY INTERDISCIPLINARY REVIEWS: COMPUTATIONAL MOLECULAR SCIENCE · NOVEMBER 2013

Impact Factor: 11.89 · DOI: 10.1002/wcms.1150

Advanced Review

Molecular interaction fields in drug discovery: recent advances and future perspectives

Anna Artese,[1] Simon Cross,[2]* Giosuè Costa,[1] Simona Distinto,[3] Lucia Parrotta,[1] Stefano Alcaro,[1] Francesco Ortuso[1] and Gabriele Cruciani[4]

Drug discovery is a highly complex and costly process, and in recent years, the pharmaceutical industry has shifted from traditional to genomics- and proteomics-based drug research strategies. The identification of druggable target sites, promising hits, and high quality leads are crucial steps in the early stages of drug discovery projects. Pharmacokinetic (PK) and drug metabolism profiling to optimize bioavailability, clearance, and toxicity are increasingly important areas to prevent costly failures in preclinical and clinical studies. The integration of a wide variety of technologies and expertise in multidisciplinary research teams combining synergistic effects between experimental and computational approaches on the selection and optimization of bioactive compounds to pass these hurdles is now commonplace, although there remain challenging areas. Molecular interaction fields (MIFs) are widely used in a range of applications to support the discovery teams, characterizing molecules according to their favorable interaction sites and therefore enabling predictions to be made about how molecules might interact. The utility of MIF-based *in silico* approaches in drug design is extremely broad, including approaches to support experimental design in hit-finding, lead-optimization, physicochemical property prediction and PK modeling, drug metabolism prediction, and toxicity.

How to cite this article:
WIREs Comput Mol Sci 2013, 00: 1–20 doi: 10.1002/wcms.1150

INTRODUCTION

The discovery and development of a new drug is an expensive and time-consuming process; therapeutic effects and hazards to health are assessed using a series of experimental and *in vivo* tests. However, usage of animal models is often subject to ethical (and financial) considerations, and therefore alternative methods are being developed to reduce the requirement of animals in testing. In particular, drug discovery has moved toward more rational strategies based on our increasing understanding of protein–ligand interactions. The combination of the available knowledge of a large number of three-dimensional protein structures with hundreds of thousands of small-molecules has attracted the attention of scientists from all over the world for the application of structure- and ligand-based drug design approaches. *In silico* methods are often implemented because of their lower cost and ability to help medicinal chemists prioritize which compounds to make; they have been shown to have made a significant contribution to the identification and development of effective drugs from new

The authors have declared no conflicts of interest in relation to this article.

*Correspondence to: simon@moldiscovery.com

[1]Dipartimento di Scienze della Salute, Università degli Studi "Magna Græcia" di Catanzaro, Campus "S. Venuta", Viale Europa, Catanzaro, Italy

[2]Molecular Discovery Ltd., Pinner, Middlesex, London, United Kingdom

[3]Dipartimento di Scienze della Vita e dell'Ambiente, Università di Cagliari, Cagliari, Italy

[4]Laboratory for Chemometrics and Cheminformatics, Chemistry Department, University of Perugia, Perugia, Italy

DOI: 10.1002/wcms.1150

chemical entities (NCEs).[1,2] One of the first studies to demonstrate the impact of computational methods on drug design, published by von Itzstein et al.,[3] highlighted the energetically favorable site for an amino or guanidine group in the active site of the influenza virus neuraminidase using the GRID program. This approach guided the design and modification of a transition-state analogue lead compound, ultimately resulting in the drug Relenza.[3]

The development of NCEs as new effective drugs is conducted under stringent conditions to ensure the therapeutic effect and the safety of the new compounds. To achieve this challenge, the benefits (therapeutic effects) and the risk (toxic effects) of the NCEs are evaluated, respectively, during the preclinical and the clinical phases of development. During the preclinical stage, the pharmacological profile and the acute toxicity of the drug candidate are assessed using *in silico*, *in vitro* methods, and animal models. For a given target (receptor, enzyme, etc.), this stage aims to identify 'hit' compounds from diverse libraries (corporate, commercial, etc.) and/or by medical observations. High-throughput screening and *in silico* evaluations are used to screen NCEs with suitable pharmacodynamic (PD) activity. The pharmacokinetic (PK) properties govern the bioavailability of the NCEs and, therefore, the correct delivery of the drug to its target site. The PK properties are represented by the processes of absorption, distribution, metabolism, and elimination (ADME) undergone by the NCEs in the organism. During this stage, 'hit' molecules presenting good ADME and physicochemical properties are identified and taken further as lead compounds. Subsequently, the lead optimization will evaluate various properties of lead analogs to propose the drug candidates. Accordingly, lead analogs are generated by producing different structural modifications around the lead's molecular scaffold. The chemical structures with the optimal potency, solubility, and ADME profile are selected as a drug candidate.

In spite of the stringent procedure and substantial financial investment of drug development, of thousands of molecules tested in the preclinical phase, only one reaches the market.[4] In one of his studies, Kennedy[5] identified the factors associated with failures during clinical assessments: poor PK properties (39%), lack of efficacy (30%), toxicity in animals (11%), and adverse effects to man (10%). Numerous computational tools, with variable success in their application, have been proposed to address ADME and potency during the early phase of drug discovery and development.[6,7] These tools range from very trivial 'rules of thumb', for example, Lipinski's rule of 5[8] to more complex and multivariate approaches, including molecular interaction field (MIF) approaches. It is encouraging that since these approaches have been introduced and applied, failures due to poor PK properties have dropped to 1%.[9]

MIFs are one of the most established and most versatile concepts in drug discovery, and were conceived to describe molecular interactions of pharmaceutical nature.[10] A MIF in fact describes the spatial variation of the interaction energy between a molecular target and a chosen probe and its calculation is mediated by software such as GRID,[11–13] which uses an energetic potential based on the total interaction energy between a target molecule and a probe (which may be an atom or a group), thus deriving distinct characteristics about the target molecule.

In ligand-based design, MIFs may be used to identify pharmacophorically similar ligands, predict bioactive alignments through pharmacophore elucidation, derive 3D-QSAR models to predict binding affinity, and to predict PK parameters, such as cell permeability and metabolism. In structure-based design, they are adopted to analyze structural features of macromolecules, and predict protein–ligand and protein–protein interactions (PPIs). MIFs have been reported to be useful in the optimization of protein–ligand interactions, areas of ligands vulnerable to metabolism, and ligand/isoform specificity.[14,15]

HIT FINDING USING MIF-BASED VIRTUAL SCREENING APPROACHES

MIFs represent a unique method for comparing molecules. They can be described by fields with the attributes that lead to their biological activity, such as the regions of positive and negative charge together with regions of high hydrophobicity in a specific shape. In 2006, Cheeseright et al. described a ligand-based method[16] for using molecular fields based on the use of a molecular mechanics force field, able to incorporate off-atom charges to obtain a more accurate representation of the electronic environment, replacing the 'grids' with the local extrema of the molecular field, thus resulting in a rapid and accurate method that was independent from changes in the used molecular coordinates.

In particular, the method generates four types of three-dimensional molecular field descriptors or 'field points' as extrema of electrostatic, steric, and hydrophobic fields. However, it is debatable as to whether the hydrophobic field is in fact a molecular

field; rather it appears to be a geometric function that locates hydrophobic points at the centroid of nonpolar regions. Nevertheless, these field points are used to define the properties necessary for a molecule to bind in a characteristic way into a specified active site. The hypothesis is that compounds showing a similar field point pattern are likely to bind at the same target site regardless of structure. The methodology to test this idea was illustrated using HIV NNRTI and thrombin ligands and validated across seven other targets. From the *in silico* comparisons of field point overlays, the experimentally observed binding poses of these ligands in their respective sites can be reproduced from pairwise comparisons. In a more recent publication, enhancements to the method and how it is applied were reported.[17] The approach has also been applied to rationalizing the activities of diverse Cholecystokinin 2 receptor antagonists,[18] and validated using the DUD dataset[19] where the approach demonstrated superior chemotype enrichment compared with a non-MIF-based docking program.

The FLAP[20] method is based on GRID MIFs in a two-step procedure. Firstly, the MIFs are condensed into discrete points representing the most favorable interactions, and all corresponding quadruplets of these points are used to generate FLAP pharmacophoric fingerprints. The fingerprints can be compared directly for increased throughput, or the individual quadruplets compared to provide different superpositions of a test molecule onto a template molecule. The superpositions are then scored in terms of their MIF field similarity; typically various combinations of hydrogen-bond donor, acceptor, hydrophobic, charged, and water MIFs are used, along with a shape field similarity. Because GRID MIFs can be produced easily for small molecules or proteins, FLAP is able to perform ligand-based screening, receptor-based screening, or protein–protein similarity calculations.

For structure-based virtual screening, the number of site points can be modified, removing those located out of the active site or adding more points to stress a particular position of interest. All the potential pharmacophores of the protein active site are calculated on the basis of the stored site points. Then, the pharmacophores of the ligands to be screened are generated using conformational sampling methods (random or systematic). Conversely to many other pharmacophoric approaches, FLAP produces a single fingerprint for each of the molecule conformations. For each conformation of each ligand under investigation, protein–ligand matches between all the possible pharmacophores of the putative active site of the protein and the pharmacophores for each ligand conformation are then calculated. The structure-based virtual screening process includes also the use of some keywords used by FLAP to filter out matches, and with the use of specific regions or interactions, constraints can be added.[10]

A recent application of FLAP was reported by Muratore et al.,[21] which identified small molecules able to effectively and specifically inhibit growth of influenza A and B viruses in cultured cells through targeting an assembly interface of the viral RNA-dependent RNA polymerase. Using an existing crystal structure of the primary protein–protein interface between the PB1 and PA subunits of the influenza A virus polymerase, three million compounds from the ZINC database were screened using FLAP and 32 molecules were selected. Finally, two compounds emerged as effective inhibitors with IC_{50} values in the low micromolar range and negligible cytotoxicity.[21]

FLAP was also applied in a recent virtual screening using a subset of the DUD (Directory of Useful Decoys) benchmarking data set containing 13 targets each with more than 15 different chemotype classes. Several ligand and receptor-based virtual screening approaches were investigated, using combinations of individual templates 2D structures of known actives, a cocrystallized ligand, a receptor structure, or a cocrystallized ligand-biased receptor structure. An excellent chemotype enrichment was achieved in both single target ligand-based and receptor-based approaches, of approximately 17-fold over random on average at a false positive rate of 1%. Moreover, if inactivity or decoy data were incorporated to train the approach, automatically the scoring function in FLAP improved, highlighting the utility of FLAP for virtual screening when either a limited or wide range of prior knowledge is available.[22] In this analysis, and also that by Cheeseright et al. mentioned above,[19] it is important to note that the virtual screening performance was analyzed in terms of the chemotype enrichment, as opposed to using only the DUD active compounds in the analysis. This demonstrates the ability of the approaches to find actives that are structurally different, an aspect that is very important and sometimes overlooked in virtual screening validation.

In other recent prospective studies, FLAP has proven successful in virtual screening approaches to identify novel openers of pancreatic K_{ATP} channels,[23] folate cycle inhibitors,[24] NorA efflux pump inhibitors,[25] and adenosine receptor subtype antagonists.[26] In this last case, the MIF-based FLAP approach demonstrated superior performance in ligand-based and structure-based approaches to other non-MIF-based methods.

LEAD OPTIMIZATION USING MIF-BASED LIGAND-BASED DESIGN

Even with the increasing availability of high quality crystallographic structures of protein targets in the past decade, ligand-based approaches are still of considerable importance in lead optimization, taking advantage of binding affinities determined by *in vitro* assays. The chances of success of a MIF-based statistical analysis approach (MFA) project strongly depend on various factors, among them the amount and quality of the binding information used as input, the statistical analysis method employed, and the quality, completeness, and balance of the MIFs used for the quantification of the interactions of the ligands with the virtual receptor.

3D-QSAR

Probably the most well-known application of MIFs is to derive three-dimensional quantitative structure–activity relationships (3D-QSAR) models by the CoMFA[27] or GRID/GOLPE[28,29] approaches for small molecules, and in this area, there are hundreds of new publications every year illustrating its continuing relevance (see Figure 1). The key advantage of 3D-QSAR over other QSAR approaches is that the results are typically straightforward and easy to interpret, enabling intuitive compound design. The price for this is paid up front; the dataset ligands must be aligned, and different alignments can affect the model significantly. In addition, care must be taken to use an appropriate statistical approach, as the number of descriptors (the MIF values at a large number of grid points) is typically large and outnumbers the measured biological value being modeled. Typically, the statistical approach used to overcome this problem is PLS (partial least square projection to latent structures), and cross-validation is also used to attempt to avoid overfitting effects and is an estimate of the predictivity of the model. The PLS model correlation coefficient is r^2 describes how well the model describes the data (1.0 being a perfect correlation), and the cross-validated r^2 (usually written as q^2) serves as a quantitative measure of the model's predicivity (1.0 being perfect, however a $q^2 > 0.3$ can be considered significant).

The first step is the modeling and the alignment of a series of compounds characterized by experimental measurements of binding or activity; subsequently, MIFs are computed for each molecule and then the values of the MIFs at the grid points are correlated with activities by means of PLS. Finally, to explain differences of activities in the series, the crucial regions around the molecules are identified, thus representing the starting point for the further design of new ligands. MIFs can be imported from a number of different sources, including GRID,[11–13] CoMFA/CoMSIA fields,[27] or quantum–mechanical (QM) electron density/electrostatic potential fields generated with a variety of QM programs, such as GAMESS,[30] GAUSSIAN,[31] JAGUAR,[32] MOLDEN,[33] and TURBOMOLE.[34] There is also the opportunity to load 3D coordinates of a dataset and compute basic force-field-based MIFs inside Open3DQSAR, a recent free, open-source tool aimed at pharmacophore exploration by high-throughput chemometric analysis of MIFs.[35] In particular, Tosco and Balle recently realized such a 3D-QSAR project aimed to overcome the template selection bottleneck by using virtually all conformers within an energetically accessible window as possible templates. Adopting this procedure on a series of nicotinic $\alpha_4\beta_2$ receptor agonists and partial agonists, they showed that, among all evaluated alignments, one compatible with pharmacophore models, site-directed mutagenesis studies, and X-ray complexes of acetylcholine binding proteins could be identified. With this aim, a 3D-QSAR model was built on each individual alignment and for each of them q^2/standard deviation of error of predictions statistics was calculated for both pIC_{50} and pEC_{50} using an external test set.[36] The

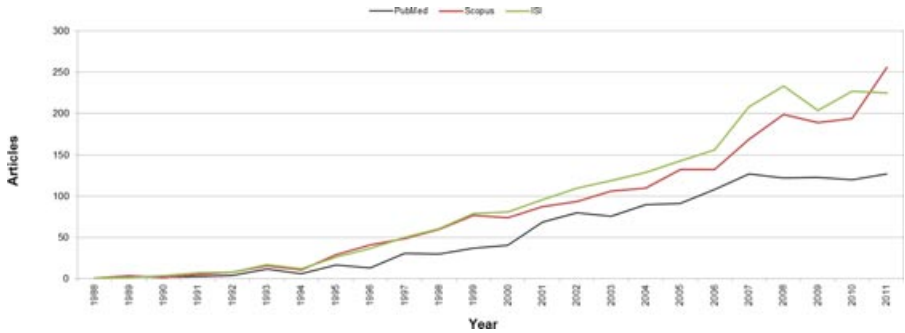

FIGURE 1 | The relevance of 3D-QSAR is illustrated by plotting the number of articles containing the 3D-QSAR keyword by year. The timeline starts at 1988 with the publication of CoMFA, numbers for 2012 include only 10 months of the year.

model showing the best predictive performance with respect to both pIC_{50} and pEC_{50}, which is consequently more likely to represent a good guess of the binding mode in the $\alpha_4\beta_2$ nicotinic receptor, is the same as the one formerly identified using GOLPE.[37]

One of the recent applications of CoMFA approach was reported by Wang et al.,[38] who developed a 3D-QSAR model of sulfonamide analogs binding a monoclonal antibody (MAbSMR) produced against sulfamerazine by using Distance Comparison (DISCOtech), CoMFA, and comparative molecular similarity indices analysis (CoMSIA). Competitive fluorescence polarization immunoassay was adopted to evaluate the affinities of the MAbSMR, for 17 sulfonamide analogs. The results demonstrated that the proposed pharmacophore model containing two hydrogen-bond acceptors, two hydrogen-bond donors, and two hydrophobic centers characterized the structural features of the sulfonamides necessary for MAbSMR binding. Removal of two outliers from the initial set of 17 sulfonamide analogs improved the predictability of the models. The 3D-QSAR models of 15 sulfonamides based on CoMFA and CoMSIA resulted in q^2_{cv} values of 0.600 and 0.523, and r^2 values of 0.995 and 0.994, respectively, thus indicating that both methods had significant predictive capability.

Another combined method based on density functional theory (DFT), molecular mechanics, and statistics as well as the comparative molecular field analysis was applied to develop 2D- and 3D-QSAR models for a novel series of ethynyl-3- quinolinecarbonitriles acting as Src kinase inhibitors.[39] The leave-one-out cross-validation q^2 values of 2D-QSAR and CoMFA models resulted 0.834 and 0.812, respectively. The predictive abilities of these models were further validated by the test set including 10 compounds, and the predicted IC_{50} values were in a good agreement with the experimental ones. Based on the established models and some design considerations, three new compounds with rather high predicted Src-inhibitory activity have been theoretically designed and used by experimenters as reference.

CoMFA and CoMSIA methods were recently adopted also for the rational design of anticancer agents. In particular, 3D-QSAR and molecular docking methods were performed on curcumin derivatives as androgen receptor antagonists, proved to be effective antiprostate cancer agents. The constructed CoMFA and CoMSIA models produced statistically significant results with the cross-validated correlation coefficients q^2 of 0.658 and 0.567, noncross-validated correlation coefficients r^2 of 0.988 and 0.978, and predicted correction coefficients r^2_{pred} of 0.715 and 0.793, respectively. Thus, a set of 30 new analogs were proposed by utilizing the results revealed in the present study, and were predicted with potential activities in the developed models.[40]

3D-QSAR models established by CoMFA and CoMSIA methods were applied also to a series of indenoisoquinolines displaying potent Topoisomerase I inhibitory activity in human renal cell carcinoma cell line SN12C.[41] Internal and external cross-validation techniques were investigated, as well as some measures taken, including region focusing, bootstrapping and the 'leave-group-out' cross-validation method. The CoMFA model predicted a q^2 value of 0.659 and an r^2 value of 0.949, indicating that electrostatic and steric properties play a significant role in potency. The best CoMSIA model, based on a combination of steric, electrostatic, and H-bond acceptor descriptors, predicted a q^2 value of 0.523 and an r^2 value of 0.902. The established predictive models proved quite reliable to efficiently guide further modification in the molecules for obtaining better anticancer drugs.

The GRID/GOLPE method[42] is a 3D-QSAR approach similar to CoMFA,[27] which for many years, has been synonymous with 3D-QSAR. CoMFA was the first technique to implement in a QSAR method, the concept that the biological activity of a ligand can be predicted by its 3D molecular fields and that a protein and a ligand interact by noncovalent reversible contacts. The 3D properties of a molecule interacting with its environment can be described by locating the molecule in a grid and calculating the interaction energies between the ligand and a probe atom at any node of the grid. In the GRID/GOLPE approach, the interaction with the probe atom is described by GRID potentials and the alignment of the molecules is required.

MIFs are extremely rich in information,[43,44] but consist of a large number of variables that are generated to describe the nonbonded interaction energies between one or more probes and each drug molecule. Much effort has been devoted to develop methods able to select only those variables of importance.[45] Moreover, generating 3D conformations and alignment for compounds used in 3D-QSAR is a complicated and time-consuming process, particularly with very flexible and large in size compounds. When the alignment problem is solved, a descriptor matrix is generated and analyzed by the PLS method. In spite of using such a powerful statistical method, spurious results can occur, thus the GOLPE, which stands for Generating Optimal Linear PLS Estimation, approach[46] was developed to identify only the meaningful variables for the prediction of the biological activity, by applying the fractional factorial design[46] and the smart region definition (SRD)[47] procedure.

The GRID/GOLPE approach has been applied to several drug targets. Among them, in 2002, Sippl[48] reported the development of a 3D-QSAR model for estrogen receptor ligands showing a significant correlation between calculated MIFs and experimentally measured binding affinity. The ligand alignment obtained from docking simulations was taken as basis for a comparative field analysis applying the GRID/GOLPE program. Using the interaction field derived with a water probe and applying the SRD procedure, a significant and robust model was obtained. To further analyze the robustness and the predictivity of the established model, several developed estrogen receptor ligands were selected as external test set. An excellent agreement between predicted and experimental binding data was obtained. Two other traditionally used prediction techniques were applied to check the performance of the receptor-based 3D-QSAR procedure. The interaction energies calculated on the basis of receptor–ligand complexes were correlated with experimentally observed affinities. Also ligand-based 3D-QSAR models were generated using the program FlexS.[49] The interaction energy-based model, as well as the ligand-based 3D-QSAR models yielded models with lower predictivity. The comparison with the interaction energy-based model and with the ligand-based 3D-QSAR models, respectively, indicated that the combination of receptor-based and 3D-QSAR methods was able to improve the quality of prediction.

More recently, Musmuca et al.[50] reported a combining computational procedure that led to the identification of novel molecular scaffolds, untested previously toward Hepatitis C virus NS5B polymerase. To select potential new NS5B nonnucleoside inhibitors, 3D-QSAR, ligand-based (LB), and structure-based (SB) alignments methods and a LB–SB virtual screening (LB–SB-VS) protocol was set up. Further, the NCI Diversity Set,[51] a database consisting of over 1,990 compounds, was virtually screened employing the LB–SB-VS strategy, and 40 molecules were selected for enzyme-based biological assays. Among the tested molecules, 10% resulted in inhibiting the NS5B RNA-dependent RNA-polymerase activity at micromolar levels. In this study, an accurate graphical analysis of GOLPE contour maps allowed to identify the most representative areas selected by the models. In fact, one of the most interesting features of a CoMFA or GRID/GOLPE 3D-QSAR analysis is the possibility of translating back the PLS coefficients assigned to each variable to the 3D positions they occupy in real space. These values can be contoured at a particular significant level and can be displayed as a grid plot of PLS coefficients. The contour coefficient maps indicate those areas in which the model has found a high correlation between the ligand–probe interaction energy and the biological activity.[50]

Although thousands of citations now exist in 3D-QSAR, its development was rather slow with the majority of new 3D-QSAR applications just extensions of CoMFA and GRID/GOLPE techniques, both using proprietary software and requiring significant user interaction. Ballante and Ragno[52] recently reported an alternative way to build 3D-QSAR models, based on an evolution of software, named 3-D QSAutogrid/R and developed to use only software freely available to academics. 3-D QSAutogrid/R was found able to improve the interpretation of the 3D-QSAR map implementing CoMFA and GRID/GOLPE by multiprobe/multiregion variable selection (MPGRS). The methodology is based on the integration of the MIFs as calculated by AutoGrid and the R statistical environment that can be easily coupled with many free graphical molecular interfaces, such as UCSF-Chimera,[53] AutoDock Tools,[54] JMol,[55] and others. Such a new 3D-QSAR procedure was applied to a data set of aligned opioid-receptor antagonists, previously described by Peng et al.[56] in a CoMFA application (LB data set), and two data sets of HCV NS5B allosteric inhibitors, as reported in Musmuca et al.[50] GRID/GOLPE analysis (SB data sets). The procedure was validated with these three case studies, improving automation and flexibility that permit the iterative generation of hundreds/thousands of 3D-QSAR models selecting the best one in a completely independent way. Furthermore, the possibility to extrapolate/merge the more informative interactions from different probe fields into a single multiprobe MIF lead to more comprehensive interpretations.[52]

As the alignment is often the bottleneck in the whole computational study, to avoid it, two commercially available software packages were recently designed and released to automatically extract the information present in MIFs in the form of numerical descriptors: Volsurf+[6] and Pentacle[57–59] (a recent improvement upon the original ALMOND software) both by Molecular Discovery.[60] In general terms, Volsurf+ descriptors are obtained from MIFs by calculating the volume or the surface of the interaction contours[6,10,61] at predefined energy values, whereas Pentacle descriptors (called GRIND) are the results of a filtering procedure based on energetic and distribution criteria and relative position of points. Thus Volsurf+ descriptors are well suited to describe some ADME properties, whereas modeling of pharmacological target-based interaction requires the use of

descriptors able to catch the specificity of the interaction as GRIND. In a recent study, Ermondi et al.[62] compared the QSAR models obtained with Volsurf+ and Pentacle for a data set of non-ATP competitive inhibitors of the Glycogen Synthase Kinase 3b (GSK-3β), involved in neurodegeneration and in particular in the Alzheimer's Disease (AD). These inhibitors were chemically related to palinurin,[63] known as a potent ATP noncompetitive inhibitor of GSK-3β. In particular, they checked whether Volsurf+ descriptors could replace GRIND in the interpretation of PD events when these latter are expressed in binary format. Results suggested not only that the simpler Volsurf+ descriptors were good enough to predict and chemically interpret the investigated phenomenon, but also a bioactive conformation of palinurin which could guide future design of ATP noncompetitive GSK-3 inhibitors.

In contrast to other 3D-QSAR methods also based on MIF mapping, such as CoMFA and CoMSIA, VolSurf models are not dependent on alignment of the molecular structures as the spatial localization and intensity of molecular interactions encoded by each MIF are condensed into 1D descriptors.[6] Quite recently, Nicolle et al.[64] have successfully developed a 3D-QSAR method based on 3D linear solvation energy analysis, combining MIFs and VolSurf descriptors, to explore the interaction forces governing the affinity of flavonoids toward a cytosolic domain of the resistance protein (BCRP/ABCG2) half transporter, that is overexpressed in breast cancer cells. The obtained results underlined the importance of hydrophobicity as a major physicochemical molecular-property for increasing inhibition of ABCG2.[64]

MIFs generated via the GRID force field were also applied to evaluate the mechanisms of retention of a series of 23 asymmetric sulfoxides.[65] Specifically, the MIFs calculated with the program GRID were used within the VolSurf program to yield original descriptors in such a way to contain information of repulsive and attractive energies among the probes and targets. This work clearly demonstrated that the MIFs are capable of discovering the nonselective interactions needed for delivering the chiral sulfoxides to the inner chiral recognition site by the amylose chiral centers.

The innovative unconventional approach COSMOsar3D, able to yield robust and predictive 3D-QSAR models, was recently reported by Klamt et al.,[66] that identified a novel set of MIFs, the local grid-based COSMO σ-profiles (LSPs), as a promising alternative to force-field-based MIFs. They demonstrated that the usage of local σ profiles in molecular field analysis inverts the role of ligands and receptors: while conventional 3D QSAR considers the virtual receptor in potential energy fields provided by the ligands, the COSMOsar3D approach corresponds to the calculation of the free energy of the ligands in a virtual free energy field provided by the receptor. Specifically, the application of this COSMOsar3D concept to the eight reference MFA datasets published by Sutherland et al.[67] highlighted a significant increase of the predictive accuracy of the resulting models compared to the standard 3D-QSAR methods. A recent alignment-free 3D QSAR study was carried out to test the antiproliferative activity of thirty-three 1,2,4,5-tetraoxane derivatives toward two human dedifferentiated cell lines by using the GRIND methodology. It was found that the pharmacophoric pattern attributed to the most potent derivatives include amido NH of the primary or secondary amide, and the acetoxy fragments at positions 7 and 12 of steroid core which are, along with the tetraoxane ring, common for all studied compounds. Independently, a simple multiple regression model obtained by using the whole-molecular properties, confirmed that the hydrophobicity and the H-bond donor properties are the main parameters influencing potency of compounds toward human cervix carcinoma (HeLa) and human malignant melanoma (FemX) cell lines.[68]

PHARMACOPHORE ELUCIDATION

Elucidation of the common pharmacophore describing the chemical features that are required but not necessarily sufficient for ligand binding at a receptor site is a key tool in modern discovery projects. Requiring only a few known ligands, pharmacophore modeling can provide insight into the three-dimensional binding mode without prior knowledge of the bioactive conformation or receptor site structure. This in turn can help rationalize the SAR of a number of ligands, and the spatial configuration of features can be used for virtual screening. Recently, Silicos NV[69] provided their freely available ligand centric pharmacophore method Pharao, which adopts three-dimensional Gaussians to reflect a molecule's pharmacophoric properties, in contrast to most methods that use conventional hard sphere models. Gaussian models show the advantage that they require far less user intervention for model creation. The MIF-based approach FLAP has been used to derive the FLAPpharm algorithm for pharmacophore elucidation.[70] With FLAPpharm, the common pharmacophore is not represented using specific features or hard sphere models, but rather pharmacophoric interaction fields (PIFs) that are the mean average MIFs across a set

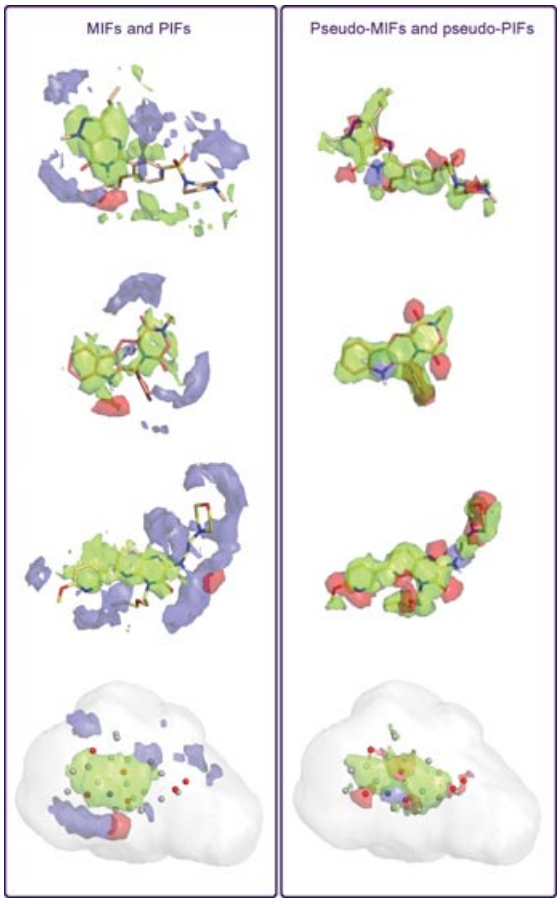

FIGURE 2 | Three PDE5 inhibitors (top of each pane) and the FLAPpharm pharmacophoric pseudomolecule (bottom of each pane) derived from a dataset of seven molecules. The GRID MIFs and PIFs are shown in the left hand pane, the pseudofields and pharmacophoric pseudofields are shown in the right hand pane. Fields are contoured transparently (hydrophobic = green, hydrogen-bond donor = blue, hydrogen-bond acceptor = red). The pharmacophoric pseudomolecule is additionally contoured transparently with a shape field in white.

of aligned ligands, and additionally pseudoPIFs and pharmacophoric points (Figure 2).

The pharmacophoric interactions are therefore weighted according to the number of ligands containing them; the resulting pharmacophoric pseudomolecule can be used as a template for virtual screening, where potential hits are ranked by their similarity to the PIFs, as opposed to classical approaches where the feature matching is a binary yes/no classification. For predicting the bioactive conformation and alignment of input ligands, FLAPpharm was validated using the PharmBench dataset containing 960 ligands and convering 81 targets.[71] According to an objective measure of success, building alignment models from the known X-ray ligand conformations was successful for 93% of the targets, whereas starting from 2D input structures gave success in 67% of the cases. A further 16% of the objectively unsuccessful cases gave subjectively good pharmacophore models (the pharmacophore may be confined to a subset of the molecules; hence, failing to align part of the molecule may not necessarily affect the pharmacaphore elucidation). When testing the performance of the software for virtual screening, FLAPpharm was used in an automatic approach to build models from the 2D structures of known actives. The screening enrichment rates were comparable to using the X-ray receptor site biased by the cocrystallized ligand, and better than using individual ligands or the pure receptor site as templates.

LEAD OPTIMIZATION USING MIF-BASED STRUCTURE-BASED DESIGN

Structure-based drug design (SBDD) has been an integral tool in the drug discovery process for more than 20 years and, with the increasing availability of experimental protein structures, its influence continues to grow. Several recent reviews[72–75] have highlighted the impact of structure on the design of compounds targeted toward the kinase, aspartyl protease, metalloprotease, and nuclear receptor families of proteins. Incorporation of structure into the design process has the powerful effect of focusing the chemical space around a certain scaffold into a space most relevant for the target of interest.

BINDING SITE CHARACTERIZATION USING MIFs

The classic MIFs application is to identify energetically favorable binding sites on a macromolecular target for probes and use these to design ligands with an improved affinity profile. MIFs can be very useful to identify areas where substituents could be added to known ligands or to design further compounds, in combination with docking and *de novo* design approaches.[76,77] SiteMap[78] is a recent approach toward characterizing binding sites in terms of hydrophobic, hydrogen-bond donor, and hydrogen-bond acceptor maps. The results are somewhat similar to the maps calculated by GRID (DRY, N1, O probes would be the corresponding GRID probes) in our experience, enabling researchers to design ligands based on these maps. The SiteMap interactions maps provide a general guide, however many differences between the approaches remain, for example, GRID enables 64 chemical probes types to map a site and more specifically enable comparison between different

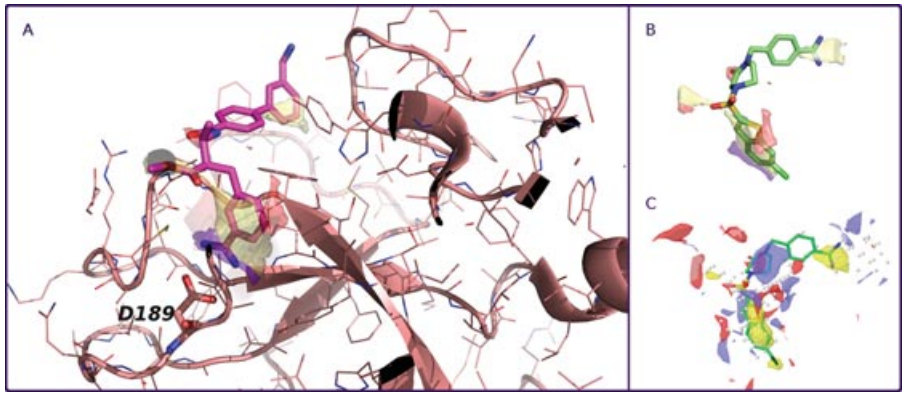

FIGURE 3 | (A) The factor Xa binding site (from 1EZQ) is shown (pink carbon atoms and ribbons) with a bound ligand (magenta carbon atoms) and Z-clipping for clarity. The GRID MIFs are contoured transparently (lipophilic = pale yellow, hydrogen-bond acceptor = red, charged hydrogen-bond donor = purple). The ligand benzamidine moiety overlaps the purple isocontour that indicates the interaction with D189 at the bottom of the S1 pocket. (B) The same MIFs with the ligand from 1NFY after protein alignment. (C) The equivalent view of (B), but with SiteMap interaction maps (lipophilic = yellow, hydrogen- bond acceptor = red, hydrogen-bond donor = blue). The lipophilic interaction maps from both approaches appear similar, however the hydrogen-bond acceptor interaction maps appear less discriminating with SiteMap, and the SiteMap hydrogen-bond donor interaction map appears to underestimate the key charge interaction with D189.

functional groups. Figure 3 shows the SiteMap and GRID maps for factor Xa, where the S1 pocket has been well characterized, typically requiring a positively charged group to interact with the aspartate D189, and bordered by hydrophobic interactions. A 'chloro binding mode' has also been reported where an additional lipophilic interaction between the chlorine atom and the cleft at the bottom of the S1 pocket is enough for chlorophenyl ligands to retain binding affinity, without the requirement of the charged group in the more typical benzamidinyl ligands. Both GRID and SiteMap highlight the chloro interaction region, whereas SiteMap appears to underestimate the aspartate charge interaction significantly. MIFs can also be applied to recognize selective regions by visual comparison of the interaction fields with different probes or with different targets, or in a systematic way by the GRID/PCA procedure.[79] Moreover, MIFs can be useful to locate the binding sites of ordered water molecules in biomacromolecules, particularly those molecules bridging protein–ligand interactions.[80]

STRUCTURE-BASED PHARMACOPHORES

Structure-based 3D pharmacophores derived solely on the interactions observed in known protein–ligand complexes may be unnecessarily restrictive. An alternative is to define pharmacophores based on an analysis of the 'hot spots' in the active site. A number of methods can in principle be used to identify these hot spots (or site points), including software such as GRID.

In this context, Ortuso et al.[81] recently developed the innovative structure-based pharmacophore approach defined GRID-based pharmacophore method (GBPM). The procedure is based on MIFs calculated with the GRID program for the ligand, receptor, and ligand–receptor complex. Using a logical combination of the field values for each MIF, the common interaction field is generated, before extracting the most relevant field nodes. These field nodes are then converted into pharmacophore features for the program Catalyst,[82] weighted by their GRID energy values.

Some interesting applications of this approach were reported. In particular, the X-linked inhibitor of apoptosis (XIAP) and the interleukin-8 dimer (IL8) were used to validate the method using known ligands and a Fit Index (FI) of these ligands to the pharmacophore. For XIAP, five known ligands were returned with a FI > 0.9 (1.0 being the maximum), hence the model was showing a high degree of recognition of the known substrates. For IL8, an NMR-derived PDB structure of the dimer was adopted, and one chain considered as the 'ligand' and returned with a FI of 0.8. Several of the residues identified by the model matched those reported by an independent study as being important for IL8 dimerization.[81]

More recently, the GBPM method was applied to 96 HIV-1 reverse transcriptase (RT) crystallographic structures to recognize the key residues involved in the p66-p51 heterodimer stabilization, as well as the RT-DNA interaction and the contacts between RT and various nonnucleoside inhibitors.[83] This study analyzed for the first time the conserved

amino acid positions of the complete RT enzyme sequence, as drug resistance is the major problem affecting the clinical efficacy of antiretroviral agents. Thus, mutagenesis data were collected from >5,500 individuals and most of the mutational observations were fully reproduced. Specifically, among the conserved residues, >90% were observed in the heterodimer analysis, whereas in the DNA and in the NNRTI recognition, 74% of the conserved residues were identified, as well as some additional invariable residues. This analysis provides further evidence of the utility of GRID-derived structure-based pharmacophores in rationalizing key molecular interactions.

Recently, the GBPM approach was applied to the NS3 protease of hepatitis C virus, and specifically in the binding-site of protease inhibitors, to highlight the most relevant residues for boceprevir target recognition. The protease residues H57, I132, S139, A156, and A157 were well identified at energy minimum threshold, emphasizing their key role in enzyme catalytic activity and stabilization. Interestingly, among all identified NS3 residues essential for boceprevir-binding by GBPM-analysis, the majority were found highly conserved among all HCV-genotypes.[84]

Lately, a hybrid model based on the Gaussian pharmacophore representation of Pharao was adapted and used in CavKA (Cavity Knowledge Acceleration), a new strategy for structure-based pharmacophore generation. CavKA interprets ligand–receptor complexes and detects interaction between ligand and binding site to derive pharmacophore models automatically. In addition, GRID MIFs can be used to weight and prioritize interacting features. By combining the smooth nature of Gaussian pharmacophores in the binding site and representing the receptor by a hard sphere excluded volume, structure-based pharmacophores can be created without any user intervention.[85]

In the same context, recently Ghersi and Sanchez[86] developed two software tools, EasyMIFs and SiteHound, that in combination are able to identify and characterize the binding sites in protein structures using an energy-based approach. EasyMIFs is a simple MIF calculator, whereas SiteHound, a post processing tool for MIFs, is able to identify interaction energy clusters corresponding to putative binding sites.[87] EasyMIFs can be used to calculate MIFs for binding site characterization, QSAR studies, selectivity analysis of protein families, pharmacophoric searching, and other applications that require MIFs. SiteHound is aimed to manipulating the output of the EasyMIFs program, as well as other programs, such as Autogrid[88] and GRID,[11–13] to predict regions on protein structures that are likely to be involved in binding ligands. The approach is based on the Q-SiteFinder algorithm,[89] but uses a different force field and clustering algorithms suited to ligands of different shapes.

DOCKING

The application of computational methods to study the formation of intermolecular complexes has been the subject of intensive research during the last decade, highlighting their importance in drug discovery projects.[90] The number of algorithms available to assess and rationalize molecular docking studies is large and ever increasing. Many algorithms share common methodologies with novel extensions, and the diversity in both their complexity and computational speed provides a plethora of techniques to deal with modern SBDD problems.[91] Several docking approaches are related to MIFs, such as Glue,[92] Glide,[93] Gold,[94] AutoDock[88] and FLAP.[20]

In particular, Glue is a docking program aimed to detect favorable modes of a ligand with respect to the protein active site using all the options and capabilities of the GRID force field. The protein cavity is mapped using several GRID runs and the resulting maps are encoded into compact files, which store the local energy minima. These minima are combined into 3D pharmacophores consisting of quartets of distance between chemical features that are automatically identified for each (macro)molecule. Then, all the accessible geometries for all the combinations of four features are calculated and stored in a fingerprint of the binding site. An iterative procedure identifies all the ways in which four atoms of the ligand could bind to the target, by pairing every atom to the nearest MIF used. Then, many orientations are eliminated because of the redundancy and steric hindrance constraints and the optimized orientations represent possible binding modes of the ligand within the site.

The first step for docking with Glide is the generation of grids that define the receptor site according to the position of the cognate ligand. For each ligand, Glide generates various conformers, places each of them in the receptor site, and minimizes them using the OPLS-AA force field with a distance-dependent dielectric. Lowest energy poses are subsequently sampled for nearby torsional minima using a Monte Carlo (MC) procedure.

Also the original procedure developed for AutoDock uses a MC simulated annealing (SA) technique for configurational exploration with a rapid energy evaluation using grid-based molecular affinity potentials. It thus combines the advantages of exploring a large search space and a robust energy evaluation. Rapid energy evaluation is achieved by

precalculating atomic affinity potentials for each atom type in the substrate molecule in the manner described by Goodford. In the AutoGrid procedure, the protein is embedded in a 3D grid and a probe atom is placed at each grid point. The energy of interaction of this single atom with the protein is assigned to the grid point. An affinity grid is calculated for each type of atom in the substrate, as well as a grid of electrostatic potential. The energetics of a particular substrate configuration is then found by trilinear interpolation of affinity values of the eight grid points surrounding each of the atoms in the substrate. The electrostatic interaction is evaluated similarly, by interpolating the values of the electrostatic potential and multiplying by the charge on the atom. SA allows an efficient exploration of the complex configurational space with multiple minima that is typical of a docking problem.

By contrast, the program Gold uses a genetic algorithm to explore the rotational flexibility of the ligand and receptor side chains. The placement of ligands is based on fitting points, which are added to the ligand and to the protein to find a match between acceptor and donor points. In addition to that, Gold uses also hydrophobic fitting points, which are mapped to CH groups of the ligand.

FLAP performs a docking-like process by describing both the protein cavity and the ligands with MIFs generated by the GRID program. It is not strictly docking, as there is no attempt to directly predict the energy of interaction between the ligand and receptor; instead the MIF field similarities are used in combination with shape, to produce docking-like poses. To mimic different types of protein–ligand interactions, it is important to include probes N1 (donor), O (acceptor), and DRY (hydrophobic) in addition to the default probe H, which describes the shape of the protein/ligand. FLAP requires two steps for docking, first where MIFs are calculated for different conformers of the ligand, and second where each of these conformers is scored according to the match between its MIFs and those of the protein cavity.

The validation of the predicted binding modes obtained after docking simulations by means of MIF calculations was recently reported by Koldsø et al., which developed a refined homology model of the human serotonin transporter (hSERT). Such a model was used for docking simulations of S- and R-citalopram, including protein-induced fit and polarization effects of the ligand, resulting in two possible binding modes for each enantiomer. To further assess the predicted binding modes from the docking simulation, the analyses were supplemented by calculations of MIFs and by computations of strain energies of the predicted binding modes compared to free citalopram. The two enantiomers were predicted to bind in the substrate binding pocket with opposite orientations of their aromatic groups. The predicted binding modes were experimentally validated using human wild type and 15 serotonin transporter mutants and 13 optically pure citalopram analogues. Crucial protein–ligand interaction points were identified confirming one binding model for each enantiomer.[95]

Recently, three docking approaches, and in particular FLAP, Glide, and Gold, were adopted by Milletti and Vulpetti to evaluate the impact of tautomerism, that is actually of special interest in studies of protein–ligand interactions.[96] Tautomer enrichment is a key step of ligand preparation prior to virtual screening. Specifically, the authors have investigated how tautomer preference in various media (water, gas phase, and crystal) compares to tautomer preference at the active site of the protein by analyzing the different possible H-bonding contacts for a set of 13 tautomeric structures. In addition, the authors explored the impact of four different protocols for the enumeration of tautomers in virtual screening by using FLAP, Glide, and Gold as docking tools on seven targets of the DUD data set. Excluding those targets in which the binding does not involve tautomeric atoms, they found that the average receiver operating characteristic curve enrichment at 10% was 0.25 (Gold), 0.24 (Glide), and 0.50 (FLAP) by considering only tautomers predicted to be unstable in water versus 0.41 (Gold), 0.56 (Glide), and 0.51 (FLAP) by limiting the enumeration process only to the predicted most stable tautomer.

Related to FLAP is the commercially available scaffold hopping method, SHOP[97] available from Molecular Discovery Ltd.[60] Such an approach employs a GRID-based method to search scaffold databases using three types of 3D-descriptors. The procedure compares the similarity of the 3D structure of a query scaffold to those in the database to find substitutes that retain the geometry, shape, and interaction patterns of the query. Conformation generation is performed on the query and the database contains precalculated conformers for each scaffold. The descriptors used are specific to the attachment or anchor points where R-groups would be attached. Distances and dihedral angles are calculated between these anchor points. Recently, Bergmann et al. performed an enrichment study to verify the ability of SHOP to find known active CDK2 scaffolds in a database. Additionally, SHOP was used for suggesting new inhibitors of p38 MAP kinase. Four p38 complexes were used to perform six scaffold searches. Several new scaffolds were suggested, and the resulting

compounds were successfully docked into the query proteins.[97]

Ligand- and structure-based methods can also be integrated to develop prediction models with wider application. In particular, a very recent publication[98] presented the first structure-based activity prediction model for benzothiadiazines against various genotypes of HCV NS5b polymerase (1a, 1b, and 4).The model was a comprehensive workflow of structure-based field template followed by guided docking. The field template was used as a prefilter and a tool to provide hits in good orientation and position. It was created based on detailed MIF analysis, while the guided docking was used as a refinement and assessment tool. The docking template was based on energy-based pharmacophore analysis. The whole procedure was formulated and tweaked for both screening [ROC of area under the curve (AUC) = 0.91] and activity prediction (r^2 of 0.8) for the genotype 1a. To widen the model scope, linear interaction energy was used as a tool for predicting activities of other genotypes based on the docked ligand poses, whereas mutation binding energy was used to investigate the effect of each amino-acid mutation in genotype 4. The model was applied for structure-based fragment hopping by screening a library designed by reaction enumeration. A top scoring hit was used to generate a focused library with better PK properties compared with the original class ligands. After that, experimental validation was carried out by the synthesis of this library and its biological evaluation, which yielded compounds that exhibit EC_{50} ranging from 1.86 to 23 μM.

ASSESSING DRUGGABILITY AND THE ROLE OF WATER

In a recent publication, Mason et al.[99] have analyzed several GPCR structures in comparison with well-studied enzyme systems to gain insights into assessing the druggability of these targets. Their analysis incorporated several methods, one of which was the MIF-based GRID approach. Using the GRID OH2 (water probe) and C1 = (aromatic carbon, lipophilic probe), they noted that the aminergic/purinergic GPCRs β1/β2-AR, A2A, D3, H1, and M2 have contiguous or clustered hydrophobic regions, but also including some water hotspots indicating hydrogen bonding sites; this pattern probably underlies the relatively high druggability of these receptors. Of particular interest is their discussion of the role of water in ligand binding, with potent ligands displacing water with a significant entropic gain. Individual water molecules interact with the receptor and water network in different ways, however, therefore predicting which water molecules are 'unhappy' could be of high importance in ligand design. Figure 4 illustrates this concept with two GPCR structures; inactive conformation adenosine A2A StaR in complex with an antagonist (PDB: 3UZA) and active conformation A2A StaR in complex with an agonist (PDB: 2YDV). The water network in the binding site of the apo structure was predicted using WaterMap[100] and SZMAP[101]; visualizing the bound ligand position for both the antagonist and agonist structures clearly shows that the ligands would displace a significant number of the 'unhappy' (red and yellow) water molecules. Almost all of these 'unhappy' water molecules lie within the GRID lipophilic isocontour, even if some of them additionally lie within the GRID water isocontour (enthalpically favorable); combining these two fields appears to provide a good estimator of the entropic gain. From a MIF perspective, such waters could be predicted and scored in a high throughput manner, providing a useful alternative approach to score predicted ligand poses or docking results.

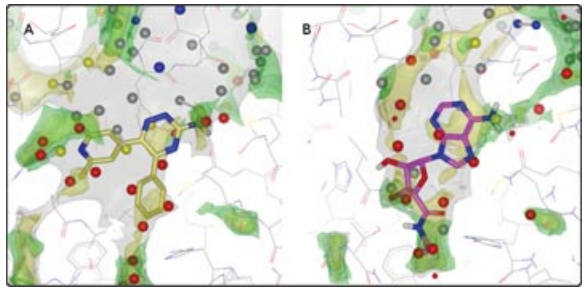

FIGURE 4 | A2A StaR in complex with an antagonist (left) and agonist (right). Explicit water molecules are shown as spheres (larger spheres from WaterMap, smaller spheres from SZMAP) and color coded according to their degree of 'unhappiness' (predicted free energy compared to bulk solvent). The 'unhappiness' color scale is red > yellow > grey > blue. GRID maps are contoured in green (OH2 water probe) and yellow (C1 = lipophilic probe). For a full description of the methodology, see Mason et al. Figure kindly provided by Andrea Bortolato. (Reproduced from Ref 99. Copyright 2012, Elsevier).

MIF-BASED PK OPTIMIZATION

As previously mentioned, major causes for failure in drug development are unsuitable PK properties of drug candidates including ADME. To obtain useful descriptors for ADME properties is not an easy task. A large number of descriptors have been developed,[102] all of which have major limitations in terms of relevance, interpretability, or speed of calculation. Calculated molecular properties from 3D molecular fields of interaction energies represent a valuable approach to correlate 3D molecular

structures with physicochemical and PD properties. The VolSurf approach[6] is able to compress the relevant information present in 3D maps into a few descriptors characterized by the simplicity of their use and interpretation. These descriptors, which refer to molecular size and shape, to hydrophilic and hydrophobic regions, and to the balance between them, can be quantitatively compared and used to build multivariate models correlating 3D molecular structures with biological responses.

In a recent application, FLAP was combined with VolSurf+ to develop a model able to predict P-glycoprotein (Pgp) inhibition and to guide compound design to modulate its impact within series of molecules. In particular, inhibitors of Pgp can be used to overcome multidrug resistance. Thus, reliable *in silico* procedures to predict Pgp inhibition are of great interest. A large and accurate literature collection yielded more than 1,200 structures; a model was then constructed using various MIF-based technologies, considering pharmacophoric features and those physicochemical properties related to membrane partitioning. High accuracy was demonstrated with two different validation sets and the information derived from the model was rationalized as a pharmacophore for competitive Pgp inhibition.[103]

The interpretation of the PK profile based on the structure is a complex task as several biological and physicochemical processes take place in parallel in the human body. The *in vivo* data are interpreted based on a number of calculated parameters, such as clearance, half-life, bioavailability, C_{max}, T_{max}, volume of distribution, and AUC. There have been various attempts to predict some of these parameters from the structure,[104,105] but in most cases, the multifactorial nature of the *in vivo* data makes it impossible to build a global model. In parallel to the ADME field, computational models are developed for specific *in vitro* assays and are based on the experimental data. The GRID MIFs have been used extensively within the ADME area to compute molecular descriptions for compounds or proteins.[11] For example, the interaction of the hydrophobic probe in GRID (the DRY probe) can be used to compute the hydrophobic surface exposed by the compound to the environment, which is related to the lipophilicity and therefore to the passive transport or to the solubility.

In a recent study, Caron et al. reported the setting-up of an *in silico* tool based on GRID/VolSurf software to predict virtual (of each conformer) chromatographic retention factors (log k30 and log kw) for Pt(II) complexes of potential antitumor activity.[106] The method is based on the parametrization of Pt(II) into GRID force field[11,12] that can be used either alone or implemented in GRID-based software (e.g., VolSurf, ALMOND, MetaSite).[9] In particular, the authors adopted VolSurf 2D descriptors, as these latter quantitatively characterize polarity and hydrophobicity, and used them to calculate virtual log P^N_{alk} of molecules in the alkane/water system. Pt(II) complexes are bound covalently to nucleotides, thus the authors used the position of water molecules extracted from a high resolution cisplatin-DNA structure deposited in the PDB (PDB code: 1I1P)[107,108] with an accurate determination of water molecules caged around ligand (cisplatin). In particular, by using GRID to calculate the MIFs for the water probe and BIOCUBE4mf[109] to select the regions satisfying energetic criteria (energy threshold), the presence of any water molecule in its observed crystallographic position was assumed to be energetically favorable. In particular, this study indicated the hydrogen bond acceptor properties of complexes as the main determinants of the steady-state volume of distribution of the five platinum drugs in clinical use and, thus, represents the first step toward the prediction of PK descriptors to be used for screening purposes in new drug design campaigns of Pt(II) antitumor candidates.

The VolSurf+ software was also adopted in another recent application with the aim to develop a reliable model for predicting Biopharmaceutics Drug Disposition Classification System (BDDCS) class, integrated with *in vitro* assays, to anticipate disposition and potential Drug–Drug Interactions (DDIs) of new molecular entities (NMEs). Specifically Broccatelli et al.[110] described a computational procedure for predicting BDDCS class from molecular structures. The model was trained on a set of 300 oral drugs, and validated on an external set of 379 oral drugs; for each molecule, a probability of BDDCS class membership was given, based on predicted extent of metabolism (EoM), Food and Drug Administration (FDA) solubility (FDAS) and their confidence scores. The accuracy in predicting FDAS was 78% in training and 77% in validation, whereas for EoM prediction, the accuracy was 82% in training and 79% in external validation. The authors further applied the BDDCS prediction model on a large set of medicinal chemistry compounds (over 30,000 chemicals). Based on this application, solubility, and not permeability, was suggested to be the major difference between NMEs and drugs. Based on BDDCS, the intestinal absorption of extensively metabolized drugs is likely to be optimal and not affected by uptake transporters. By using predicted EoM, it could be possible to identify molecules for which passive permeability testing is not necessary. For these molecules, the modulation of metabolic enzyme activity could result

in severe adverse drug reactions. In contrast, NMEs predicted to be poorly metabolized should target uptake transporters, to optimize their intestinal absorption. This awareness could allow scientists to forecast DDIs with therapeutics that induce or inhibit uptake transporters. Another recent study by Brocatelli et al. used MIF-based approaches, reinforced by the BDDCS method, to suggest that high Torsade de Pointes risk stems from an interplay between hERG inhibition, EoM, active transport, and solubility.[111]

MIF-BASED DRUG METABOLISM PREDICTION

In the drug discovery process, the study of metabolism of NCEs is carried out to clarify several aspects, such as: the rate and site of metabolism, enzymes and tissues selectivity, and enzyme inhibition and induction responsible for drug interactions. Metabolic transformations are frequently related to the incidence of toxic effects that may result from the emergence of reactive species, the systemic accumulation of metabolites, or by induction of metabolic pathways. Recently, many publications have shown computational methods trying to address metabolic issues. GRID-based models have contributed in the field of cytochrome inhibition,[112–115] site of metabolism prediction,[116] selectivity analysis,[117,118] selective site of metabolism prediction and metabolic stability.[119] A new technology to predict P450 inhibition, metabolic stability, and isoform selectivity has also been briefly introduced,[120] using MIF approaches in combination with statistical methods, and high quality metabolism data.

In particular, two main factors determine the site of metabolism: the chemical reactivity and the preferred orientation of the compound inside the cytochrome cavity. Thus, MetaSite[121] has been developed to consider at the same time the substrate–cytochrome interaction and the chemical reactivity of the compounds toward oxidation. The recognition part compares the interaction profile of the enzyme based on GRID-MIFs and different conformations of the potential ligands. The reactivity part comes from precomputed reactivity values of fragments that are recognized in the structure under consideration. The prediction rate for the site of metabolism for five cytochromes has been validated using more than 900 metabolic reactions.

By contrast, QMBO[122] is able to estimate hydrogen abstraction energies based on bond order. The method, recently applied onto 81 molecules and able to predict 84%,[122] relates the reactivity of each hydrogen atom to the strength of its covalent bond. Using a wave function generated from DFT, bond orders for all C–H bonds in a substrate are calculated, and then normalized. Bond strength is correlated to deviations from average bond orders. Corrections are made for buried hydrogen atoms through scaling by a factor that is a function of the solvent accessible surface area of the hydrogen atom.

Of particular interest is the combination of predictive approaches and experimental data analysis, which has been published by Bonn et al.[123] Typically, metabolite identification is performed during the lead optimization process, on a relatively small number of compounds, given the amount of material and effort required to interpret the data. Identifying reactive metabolites, therefore enabling safer compounds to be selected, and additionally would allowing chemists to design metabolically stable series at an earlier point in the process, would be hugely beneficial. The

BOX 1: FUTURE PERSPECTIVE—DISRUPTING PPIs USING MIFs

PPIs are central to most biological processes, and represent an important target for therapeutic agents. Only in the last decade has opinion changed that small molecule drugs could compete with the much larger native protein partner. Several PPI inhibitors have been identified and they are typically characterized as larger and more hydrophobic compared with conventional drugs. The interaction site itself is also naturally much larger and hydrophobic; it is also much flatter than a typical enzyme or receptor site. Conventional structure-based *in silico* methods for identifying ligands may be less appropriate to find PPI inhibitors; it is likely that they have been parameterized using enzyme– or receptor–ligand complexes, they may underestimate the hydrophobic contribution, and they may not be able to use the hydrophobic regions to generate binding poses to start with (many docking algorithms place poses by matching hydrogen-bonding interactions). MIFs are flexible in application, and two of these are particularly relevant to disrupting PPI. Firstly, the PPI interface from a known complex can be characterized in a straightforward manner using the GBPM approach (described in the text). Secondly, structure-based pose prediction is possible using several methods, including the FLAP approach. Figure 5 illustrates a proof-of-concept application to the p53-HDM2 PPI cancer therapy target. The interaction site MIFs are identified and subsequently used with FLAP to 'dock' a known PPI inhibitor. The experimental pose is predicted almost perfectly and with a good MIF similarity score; using this MIF approach may therefore help in virtual screening and structure-based design in this challenging area to find other PPI inhibitors.

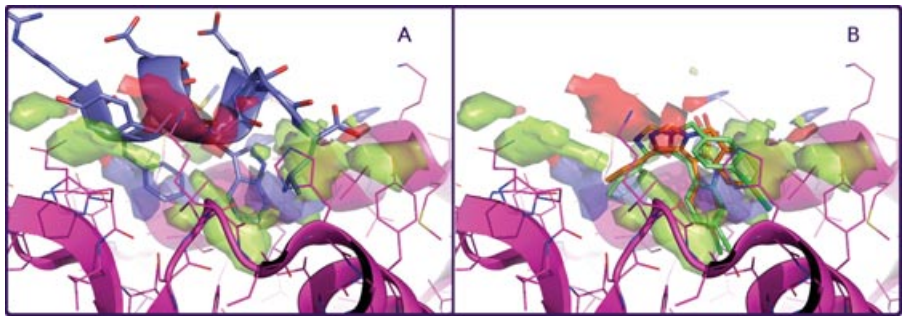

FIGURE 5 | (A) The interaction site between HDM2 (magenta carbon atoms and ribbon) and an optimized p53 peptide (pale blue carbon atoms and ribbon) is shown with the GBPM GRID MIFs contoured transparently (hydrophobic = green, hydrogen-bond donor = blue, hydrogen-bond acceptor = red). (B) The same HDM2 interaction site is shown with a small molecule PPI inhibitor from 3LBK after protein alignment (orange carbon atoms), and the same inhibitor 'docked' using FLAP (pale green carbon atoms). The predicted pose is 1.1 Å rmsd from the X-ray position.

Mass-MetaSite software[124] incorporates this approach and in validation gave over 80% success in automatically assigning metabolites from LC–MS/MS experimental data. The approach is linked to the MIF-based MetaSite approach described above; ambiguous Markush fragments identified from the experimental data can be refined to the atomic position of the biotransformation using the predictive algorithm. Manual interpretation of a single incubation is typically several hours; using an automated approach that provides results within minutes removes this bottleneck in the process; making high-throughput metabolite identification a reality.

CONCLUSIONS

With drug discovery facing increasing challenges, and increasing amounts of experimental data from disciplines such as genomics and proteomics, it is more important than ever to develop, optimize, and use computational tools to help support the experimental design process. Now that GPCR receptor structures are becoming available, structure-based design approaches will play an important role in aiding the design of new therapeutic agents targeting this class, and structure-based MIF approaches have already been used to find novel ligands against a number of these. New applications of MIFs toward the classical techniques of virtual screening, pharmacophore elucidation, and docking have been reported, improving upon previous methods in an incremental fashion. There are also, however, several areas where MIF-based approaches may lead to a 'step-jump' beyond current capabilities. New methods for predicting the role of water in ligand binding are being developed, using MIFs, and this could well improve the scoring of predicted ligands by more explicit consideration of the entropic effect. While the *common reference framework* approach to describing and comparing MIFs has primarily been applied to ligand-based approaches and ligand–receptor similarity, it is easy to see how this could be extended to compare protein pockets and PPI sites. This would allow the prediction of off-target effects, identify druggable sites, target PPIs, and potentially predict protein interaction networks. From a PK perspective, one of the biggest 'step-jumps' has already occurred using MIF-based approaches, with high throughput metabolite identification now a reality. This therefore allows this critical discovery step to be performed much earlier in the process, and also allows experts to more fully characterize compounds across a panel of isoforms and species to enable better analysis. MIF-based approaches to predicting CYP inhibition, interaction with transporters, and risk of Torsade de Pointes are helping to improve toxicity prediction. MIFs have long proven their utility in a range of applications; however, it is clear that in some areas they are quite literally defining the field.

REFERENCES

1. Drayer JI, Burns JP. *Burger's Medicinal Chemistry and Drug Discovery*. New York: Wiley; 1995.
2. Lesko LJ, Rowland M, Peck CC, Blaschke TF, Breimer D, de Jong HJ, Grahnen A, Kuhlmann JJ, Stewart B. Optimizing the science of drug development: opportunities for better candidate selection and accelerated evaluation in humans. *Eur J Pharm Sci* 2000, 10:9–14.
3. von Itzstein M, Wu WY, Kok GB, Pegg MS, Dyason JC, Jin B, Van Phan T, Smythe ML, White HF, Oliver

SW, et al. Rational design of potent sialidase-based inhibitors of influenza virus replication. *Nature* 1993, 363:418–423.

4. DiMasi JA. Success rates for new drugs entering clinical testing in the United States. *Clin Pharmacol Ther* 1995, 58:1–14.
5. Kennedy T. Managing the drug discovery/development interface. *Drug Disc Today* 1997, 2:436–444.
6. Cruciani G, Crivori P, Carrupt P-A, Testa B. Molecular fields in quantitative structure permeation relationships: the VolSurf approach. *J Mol Struct: Theochem* 2000, 503:17–30.
7. Lombardo F, Gifford E, Shalaeva MY. In silico ADME prediction: data, models, facts and myths. Mini Rev Med Chem 8:861–875; Hansch C, Leo A, Mekapati SB, Kurup A. QSAR and ADME. *Bioorg Med Chem* 2004, 12:3391–3400.
8. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Del Rev* 1997, 23:3–25.
9. Arrowsmith J. Trial watch: phase II failures: 2008–2010. *Nat Rev Drug Discov* 2011, 10:328–329.
10. Cruciani G. *Molecular Interaction Fields—Applications in Drug Discovery and ADME Prediction*. Zurich: Wiley-VCH; 2005.
11. Goodford PJ. A computational procedure for determining energetically favorable binding sites on biological important macromolecules. *J Med Chem* 1985, 28:849–857.
12. Wade RC, Clark KJ, Goodford PJ. Further development of hydrogen bond functions for use in determining energetically favorable binding sites on molecules of known structure. 1. Ligand probe groups with the ability to form two hydrogen bonds. *J Med Chem* 1993, 36:140–147.
13. Wade RC, Goodford PJ. Further development of hydrogen bond functions for use in determining energetically favorable binding sites on molecules of known structure. 2. Ligand probe groups with the ability to form more than two hydrogen bonds. *J Med Chem* 1993, 36:148–156.
14. Kirchmair J, Williamson MJ, Tyzack JD, Tan L, Bond PJ, Bender A, Glen RC. Computational prediction of metabolism: sites, products, SAR, P450 enzyme dynamics, and mechanisms. *J Chem Inf Model* 2012, 52:617–648.
15. Cross S, Cruciani G. Molecular fields in drug discovery: getting old or reaching maturity? *Drug Disc Today* 2010, 15:23–32.
16. Cheeseright T, Mackey M, Rose S, Vinter A. Molecular field extrema as descriptors of biological activity: definition and validation. *J Chem Inf Model* 2006, 46:665–676.
17. Cheeseright T, Mackey M, Scoffin, RA. High content pharmacophores from molecular fields: a biologically relevant method for comparing and understanding ligands. *Curr Comput Aided Drug Des* 2011, 7:190–20.
18. Low CMR, Vinter JG. Rationalizing the activities of diverse cholecystokinin 2 receptor antagonists using molecular field points. *J Med Chem* 2008, 51:565–573.
19. Cheeseright TJ, Mackey MD, Melville JL, Vinter JG. FieldScreen: virtual screening using molecular fields. Application to the DUD data set. *J Chem Inf Model* 2008, 48:2108–2117.
20. Baroni M, Cruciani G, Sciabola S, Perruccio F, Mason JS. A common reference framework for analyzing/comparing proteins and ligands. Fingerprints for ligands and proteins (FLAP): theory and application. *J Chem Inf Model* 2007, 47:279–294.
21. Muratore G, Goracci L, Mercorelli B, Foeglein A, Digard P, Cruciani G, Palù G, Loregian A. Small molecule inhibitors of influenza A and B viruses that act by disrupting subunit interactions of the viral polymerase. *Proc Natl Acad Sci USA* 2012, 109:6247–6252.
22. Cross S, Baroni M, Carosati E, Benedetti P, Clementi S. FLAP: GRID molecular interaction fields in virtual screening. Validation using the DUD data set. *J Chem Inf Model* 2010, 50:1442–1450.
23. Carosati E, Mannhold R, Wahl P, Hansen JB, Fremming T, Zamora I, Cianchetta G, Baroni M. Virtual screening for novel openers of pancreatic KATP channels. *J Med Chem* 2007, 50:2117– 2126.
24. Carosati E, Sforna G, Pippi M, Marverti G, Ligabue A, Guerrieri D, Piras S, Guaitoli G, Luciani R, Costi MP, Cruciani G. Ligand-based virtual screening and ADME-tox guided approach to identify triazolo-quinoxalines as folate cycle inhibitors. *Bioorg Med Chem* 2010, 18:7773–7785.
25. Brincat JP, Carosati E, Sabatini S, Manfroni G, Fravolini A, Raygada JL, Patel D, Kaatz GW, Cruciani G. Discovery of novel inhibitors of the NorA multidrug transporter of *Staphylococcus aureus*. *J Med Chem* 2011, 54:354–365.
26. Sirci F, Goracci L, Rodriguez D, van Muijlwijk-Koezen J, Gutierrez-de-Teran H, Mannhold R. Ligand-, structure- and pharmacophore-based molecular fingerprints; a case study on adenosine A1, A2A, A2B, and A3 receptor antagonists. *J Comput Aided Mol Des* 2012, 26:1247–1266.
27. Cramer RD, Patterson DE, Bunce JD. Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. *J Am Chem Soc* 1988, 110:5959–5967.
28. Pastor M, Cruciani G, Watson KA. A strategy for the incorporation of water molecules present in a ligand binding site into a three-dimensional quantitative

structure-activity relationship analysis. *J Med Chem* 1997, 40:4089–4102.

29. Ortiz AR, Pastor M, Palomer A, Cruciani G, Gago F, Wade RC. Reliability of comparative molecular field analysis models: effects of data scaling and variable selection using a set of human synovial fluid phospholipase A2 inhibitors. *J Med Chem* 1997, 40:1136–1148.
30. Schmidt MW, Baldridge KK, Boatz JA, Elbert ST, Gordon MS, Jensen JH, Koseki S, Matsunaga N, Nguyen KA, Su S, et al. General atomic and molecular electronic structure system. *J Comput Chem* 1993, 14:1347–1363.
31. Gaussian 03, revision C.02. Wallingford, CT, USA; 2004. http://www.gaussian.com. Accessed December 24, 2009.
32. Jaguar version 7.6. Schrödinger, LLC, New York, NY, USA; 2009. http://www.schrodinger.com. Accessed December 24, 2009.
33. Schaftenaar G, Noordik JH. Molden: a pre- and post-processing program for molecular and electronic structures. *J Comput Aided Mol Des* 2000, 14:123–134.
34. TURBOMOLE V6.0 a development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 1989–2007, TURBOMOLE GmbH, since 2007. http://www.turbomole.com. Accessed December 24, 2009.
35. Tosco P, Balle T. Open3DQSAR: a new open-source pharmacophore explorer based on chemometric analysis of molecular interaction fields. *J Mol Model* 2011, 17:201–208.
36. Tosco P, Balle T. A 3D-QSAR-driven approach to binding mode and affinity prediction. *J Chem Inf Model* 2012, 52:302–307.
37. Tosco P, Ahring PK, Dyhring T, Peters D, Harpsøe K, Liljefors T, Balle T. Complementary three-dimensional quantitative structure-activity relationship modeling of binding affinity and functional potency: a study on $\alpha 4\beta 2$ nicotinic ligands. *J Med Chem* 2009, 52:2311–2316.
38. Wang Z, Kai Z, Beier RC, Shen J, Yang X. Investigation of antigen–antibody interactions of sulfonamides with a monoclonal antibody in a fluorescence polarization immunoassay using 3D-QSAR models. *Int J Mol Sci* 2012, 13:6334–6351.
39. Fang DQ, Wu WJ, Zhang R, Zeng GH, Zheng KC. Theoretical studies of QSAR and molecular design on a novel series of ethynyl-3-quinolinecarbonitriles as SRC inhibitors. *Chem Biol Drug Des* 2012, 80:134–147.
40. Xu G, Chu Y, Jiang N, Yang J, Li F. The three dimensional quantitative structure activity relationships (3D-QSAR) and docking studies of curcumin derivatives as androgen receptor antagonists. *Int J Mol Sci* 2012, 13:6138–6155.
41. Zhi Y, Yang J, Tian S, Yuan F, Liu Y, Zhang Y, Sun P, Song B, Chen Z. Quantitative structure–activity relationship studies on indenoisoquinoline topoisomerase i inhibitors as anticancer agents in human renal cell carcinoma cell line SN12C. *Int J Mol Sci* 2012, 13:6009–6025.
42. Nilsson J, Wikström H, Smilde A, Glase S, Pugsley T, Cruciani G, Pastor M, Clementi S. GRID/GOLPE 3D quantitative structure–activity relationship study on a set of benzamides and naphthamides, with affinity for the dopamine D3 receptor subtype. *J Med Chem* 1997, 40:833–840.
43. Caron G, Nurisso A, Ermondi G. How to extend the use of grid-based interaction energy maps from chemistry to biotopics. *Chem MedChem* 2009, 4:29–36.
44. Caron G, Ermondi G. Calculating virtual log P in the alkane/water system (log PN^{alk}) and its derived parameters $\Delta logPN^{oct\text{-}alk}$ and log DpH^{alk}. *J Med Chem* 2005, 48:3269–3279.
45. Cruciani G. Comparative molecular fuel analysis using GRID force-field and GOLPE variable selection methods in a study of inhibitors of glycogen b. *J Med Chem* 1994, 37:2589–2601.
46. Baroni M, Constantino G, Cruciani G, Riganelli D, Valigli R, Clementi S. Generating optimal linear PLS estimation (GOLPE): an advanced chemometric tool for handling 3D-QSAR problems. *Quant Struct–Act Relat* 1993, 12:9–20.
47. Pastor M, Cruciani G, Clementi S. Smart region definition: a new way to improve the predictive ability and interpretability of three-dimensional quantitative structure-activity relationships. *J Med Chem* 1997, 40:1455–1464.
48. Sippl W. Binding affinity prediction of novel estrogen receptor ligands using receptor-based 3D-QSAR methods. *Bioorg Med Chem* 2002, 10:3741–3755.
49. Lemmen C, Lengauer T, Klebe G. FlexS: a method for fast flexible ligand superposition. *J Med Chem* 1998, 41:4502–4520.
50. Musmuca I, Caroli A, Mai A, Kaushik-Basu N, Arora P, Ragno R. Combining 3-D quantitative structure–activity relationship with ligand based and structure based alignment procedures for in silico screening of new hepatitis C virus NS5B polymerase inhibitors. *J Chem Inf Model* 2010, 50:662–676.
51. http://dtp.nci.nih.gov. Accessed June 2009.
52. Ballante F, Ragno R. 3-D QSAutogrid/R: an alternative procedure to build 3D-QSAR models. Methodologies and applications. *J Chem Inf Model* 2012, 52:1674–1685.
53. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. UCSF Chimera-a visualization system for exploratory

research and analysis. *J Comput Chem* 2004, 25:1605–1612.

54. Morris GM, Goodsell DS, Halliday RS, Huey R, Hart WE, Belew RK, Olson AJ. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J Comput Chem* 1998, 19:1639–1662.
55. Willighagen EL. Processing CML conventions in Java. *Internet J Chem* 2001, 4:4–13.
56. Peng Y, Keenan SM, Zhang Q, Kholodovych V, Welsh WJ. 3D-QSAR comparative molecular field analysis on opioid receptor antagonists: pooling data from different studies. *J Med Chem* 2005, 48:1620–1629.
57. Pastor M, Cruciani G, McLay I, Pickett S, Clementi S. GRid-Independent descriptors (GRIND): a novel class of alignment-independent three-dimensional molecular descriptors. *J Med Chem* 2000, 43:3233–3243.
58. Pastor M. Development and validation of AMANDA, a new algorithm for selecting highly relevant regions in molecular interaction fields. *J Chem Inf Mod* 2008, 48:1813–1823.
59. Durán A, Zamora I, Pastor M. Suitability of GRIND-based principal properties for the description of molecular similarity and ligand-based virtual screening. *J Chem Inf Mod* 2009, 49:2129–2138.
60. Molecular Discovery Ltd., UK. http://www.moldiscovery.com. Accessed December 2012.
61. Crivori P, Cruciani G, Carrupt PA, Testa B. Predicting blood–brain barrier permeation from three-dimensional molecular structure. *J Med Chem* 2000, 110:2204–2216.
62. Ermondi G, Caron G, Pintos IG, Gerbaldo M, Pérez M, Pérez DI, Gándara Z, Martínez A, Gómez G, Fall Y. An application of two MIFs-based tools (Volsurf+ and Pentacle) to binary QSAR: the case of a palinurin-related data set of non-ATP competitive glycogen synthase kinase 3β (GSK-3β) inhibitors. *Eur J Med Chem* 2011, 46:860–869.
63. Alonso D, Martinez A. *Glycogen Synthase Kinase 3 (GSK-3) and Its Inhibitors: Drug Discovery and Developments*. New Jersey: Wiley-Interscience; 2006, 257–280.
64. Nicolle E, Boccard J, Guilet D, Dijoux-Franca MG, Zelefac F, Macalou S, Grosselin J, Schmidt J, Carrupt PA, Di Pietro A, Boumendjel A. Breast cancer resistance protein (BCRP/ABCG2): new inhibitors and QSAR studies by a 3D linear solvation energy approach. *Eur J Pharm Sci* 2009, 38:39–46.
65. Montanari ML, Cass QB, Leitão A, Andricopulo AD, Montanari CA. The role of molecular interaction fields on enantioselective and nonselective separation of chiral sulfoxides. *J Chromatogr A* 2006, 1121:64–75.
66. Klamt A, Thormann M, Wichmann K, Tosco P. COSMOsar3D: molecular field analysis based on local COSMO σ-profiles. *J Chem Inf Model* 2012, 52:2157–2164.
67. Sutherland JJ, O'Brien LA, Weaver DF. A comparison of methods for modeling quantitative structure-activity relationships. *J Med Chem* 2004, 47:5541–5554.
68. Cvijetic IN, Zizak ZP, Stanojkovic TP, Juranic ZD, Terzic N, Opsenica IM, Opsenica DM, Juranic IO, Drakulic BJ. An alignment independent 3D QSAR study of the antiproliferative activity of 1,2,4,5-tetraoxanes. *Eur J Med Chem* 2010, 45:4570–4577.
69. Pharao 3.0.3, Silicos NV. http://www.silicos.be. Accessed December 2012.
70. Cross S, Baroni M, Goracci L, Cruciani G. GRID-based three-dimensional pharmacophores I: FLAPpharm, a novel approach for pharmacophore elucidation. *J Chem Inf Model* 2012, 52:2587–2598.
71. Cross S, Ortuso F, Baroni M, Costa G, Distinto S, Moraca F, Alcaro S, Cruciani G. GRID-based three-dimensional pharmacophores II: PharmBench, a benchmark data set for evaluating pharmacophore elucidation methods. *J Chem Inf Model* 2012, 52:2599–2608.
72. Lewis RA. Computer-aided drug design 2005–2007. *Chem Modell* 2008, 5:51–66.
73. Lang PT, Aynechi T, Moustakas D, Shoichet B, Kuntz ID, Brooijmans N, Oshiro CM. *Molecular Docking and Structure-Based Design In: Drug Discovery Research: New Frontiers in the Post-Genomic Era*. New Jersey: Wiley; 2007, 3–23.
74. Abraham DJ. Structure-based drug design—a historical perspective and the future. *Compr Med Chem II* 2006, 4:65–86.
75. Lange G. Structure-based drug design—the use of protein structure in drug discovery. *Compr Med Chem II* 2006, 4:597–650.
76. Pisabarro MT, Ortiz AR, Palomer A, Cabre F, Garcia L, Wade RC, Gago F, Mauleon D, Carganico G. Rational modification of human synovial fluid phospholipase A2 inhibitors. *J Med Chem* 1994, 37:337–341.
77. Bitomsky W, Wade RC. Docking of glycosaminoglycans to heparin-binding proteins: validation for aFGF, bFGF, and antithrombin and application to IL-8. *J Am Chem Soc* 1999, 121:3004–3013.
78. Halgren T. New method for fast and accurate binding-site identification and analysis. *Chem Biol Drug Des* 2007, 69:146–148.
79. Kastenholz MA, Pastor M, Cruciani G, Haaksma EE, Fox T. GRID/CPCA: a new computational tool to design selective ligands. *J Med Chem* 2000, 43:3033–3044.

80. Fornabaio M, Spyrakis F, Mozzarelli A, Cozzini P, Abraham DJ, Kellogg GE. Simple, intuitive calculations of free energy of binding for protein-ligand complexes. 3. The free energy contribution of structural water molecules in HIV-1 protease complexes. *J Med Chem* 2004, 47:4507–4516.
81. Ortuso F, Langer T, Alcaro S. GBPM: GRID-based pharmacophore model: concept and application studies to protein–protein recognition. *Bioinformatics* 2006, 22:1449–1455.
82. Accelrys Inc., San Diego. http://www.accelrys.com/. Accessed December 2012.
83. Alcaro S, Artese A, Ceccherini-Silberstein F, Chiarella V, Dimonte S, Ortuso F, Perno CF. Computational analysis of human immunodeficiency virus (HIV) type-1 reverse transcriptase crystallographic models based on significant conserved residues found in highly active antiretroviral therapy (HAART)-treated patients. *Curr Med Chem* 2010, 17:290–308.
84. Cento V, Mirabelli C, Salpini R, Dimonte S, Artese A, Costa G, Mercurio F, Svicher V, Parrotta L, Bertoli A, et al. HCV genotypes are differently prone to the development of resistance to linear and macrocyclic protease inhibitors. *PLoS ONE* 2012, 7:e39652.
85. Koelling F, Baumann K. CavKA^{HYBRID}—between hard spheres and Gaussians. *J Cheminf* 2012, 4(Suppl 1):P20. From 7th German Conference on Chemoinformatics: 25 CIC-Workshop Goslar, Germany, 6–8 November 2011.
86. Ghersi D, Sanchez R. Beyond structural genomics: computational approaches for the identification of ligand binding sites in protein structures. *J Struct Funct Genomics* 2011, 12:109–117.
87. Ghersi D, Sanchez R. EasyMIFS and SiteHound: a toolkit for the identification of ligand-binding sites in protein structures. *Bioinformatics* 2009, 25:3185–3186.
88. Morris GM, Goodsell DS, Halliday RS, Huey R, Hart WE, Belew RK, Olson AJ. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J Comput Chem* 1998, 19:1639–1662.
89. Laurie AT, Jackson RM. Q-SiteFinder: an energy-based method for the prediction of protein-ligand binding sites. *Bioinformatics* 2005, 21:1908–1916.
90. Blaney JM, Dixon JS. A good ligand is hard to find: automated docking methods. *Perspect Drug Discov* 1993, 1:301–319.
91. Abagyan R, Totrov M. High-throughput docking for lead generation. *Curr Opin Chem Biol* 2001, 5:375–382.
92. Sciabola S, Baroni M, Carosati E, Cruciani G. Recent improvements in the GRID force field. 1. The docking procedure GLUE. Poster Presented at 15th Eur. Symp. QSAR & Molecular Modelling, Istanbul, Turkey; 2004.
93. Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, Mainz DT, Repasky MP, Knoll EH, Shelley M, Perry JK, et al. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J Med Chem* 2004, 47:1739–1749.
94. Jones G, Willett P, Glen RC. Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation. *J Mol Biol* 1995, 245:43–53.
95. Koldsø H, Severinsen K, Tran TT, Celik L, Jensen HH, Wiborg O, Schiøtt B, Sinning S. The two enantiomers of citalopram bind to the human serotonin transporter in reversed orientations. *J Am Chem Soc* 2010, 132:1311–1322.
96. Milletti F, Vulpetti A. Tautomer preference in PDB complexes and its impact on structure-based drug discovery. *J Chem Inf Model* 2010, 50:1062–1074.
97. Bergmann R, Liljefors T, Sørensen MD, Zamora I. SHOP: receptor-based Scaffold HOPping by GRID-based similarity searches. *J Chem Inf Model* 2009, 49:658–669.
98. Ismail MA, Abou El Ella DA, Abouzid KA, Mahmoud AH. Integrated structure-based activity prediction model of benzothiadiazines on various genotypes of HCV NS5b polymerase (1a, 1b and 4) and its application in the discovery of new derivatives. *Bioorg Med Chem* 2012, 20:2455–2478.
99. Mason S, Bortolato A, Congreve M, Marshall, FH. New insights from structural biology into the druggability of G protein-coupled receptors. *Trends Pharm Sci* 2012, 33:249–260.
100. Schrodinger LLC. New York. http://www.schrodinger.com. Accessed December 2012.
101. OpenEye Scientific Software, Inc. New Mexico. http://www.eyesopen.com. Accessed December 2012.
102. Todeschini R, Consonni V. Methods and Priciples in Medicinal Chemistry. *VCH, Weinheim*; 2000, 11.
103. Broccatelli F, Carosati E, Neri A, Frosini M, Goracci L, Oprea TI, Cruciani G. A novel approach for predicting P-glycoprotein (ABCB1) inhibition using molecular interaction fields. *J Med Chem* 2011, 54:1740–1751.
104. Yoshida F, Topliss JG. QSAR model for drug human oral bioavailability. *J Med Chem* 2000, 43:4723.
105. Zamora I, Oprea TI, Ungell AL. Prediction of oral drug permeability in rational approaches to drug design. *Prous Science*; 2001, 271–280.
106. Caron G, Ravera M, Ermondi G. Molecular interaction fields (MIFs) to predict lipophilicity and ADME profile of antitumor Pt(II) complexes. *Pharm Res* 2011, 28:640–646.
107. Coste F, Malinge JM, Serre L, Shepard W, Roth M, Leng M, Zelwar C. Crystal structure of a double-stranded DNA containing a cisplatin interstrand cross-link at 1.63Å resolution: hydration at the platinated site. *Nucleic Acids Res* 1999, 27:1837–1846.

108. Coste F, Shepard W, Zelwer C. Description of ordered solvent molecules in a platinated decanucleotide duplex refined at 1.6Å resolution against experimental MAD phases. *Acta Crystallogr Sect D* 2002, 58:431–440.
109. Caron G, Nurisso A, Ermondi G. How to extend the use of grid-based interaction energy maps from chemistry to biotopics. *Chem MedChem* 2009, 4:29–36.
110. Broccatelli F, Cruciani G, Benet LZ, Oprea TI. BDDCS class prediction for new molecular entities. *Mol Pharm* 2012, 9:570–580.
111. Broccatelli F, Manhold R, Moriconi A, Giuli S, Carosati E. QSAR modeling and data mining link torsades de pointes risk to the interplay of extent of metabolism, active transport, and hERG liability. *Mol Pharm* 2012, 9:2290–2301.
112. Afzelius L, Zamora I, Ridderström M, Andersson TB, Karlén A, Masimirembwa CM. Competitive CYP2C9 inhibitors: enzyme inhibition studies, protein homology modeling, and three-dimensional quantitative structure–activity relationship analysis. *Mol Pharmacol* 2001, 59:909–919.
113. Ekin S, Bravi G. Three- and four-dimensional-quantitative structure activity relationship (3D/4D QSAR) analyses of CYP2C9 inhibitors. *Drug Metab Disp* 2000, 28:994–1002.
114. Afzelius L, Andersson TB, Karlén A, Masimirembwa CM, Zamora I. Discriminant and quantitative PLS analysis of competitive CYP2C9 inhibitors versus non-inhibitors using alignment independent GRIND descriptors. *J Comput Aided Mol Des* 2002, 16:443–458.
115. Afzelius L, Andersson TB, Karlén A, Masimirembwa CM, Zamora I. Conformer- and alignment-independent model for predicting structurally diverse competitive CYP2C9 inhibitors. *J Med Chem* 2004, 47:907–914.
116. Zamora I, Afzelius L, Cruciani G. Predicting drug metabolism: a site of metabolism prediction tool applied to the cytochrome P450 2C9. *J Med Chem* 2003, 46:2313–2324.
117. Ridderström M, Zamora I, Fjaströ m O, Andersson TB. Analysis of selective regions in the active site of human CYP 2C8, 2C9, 2C18 and 2C19 homology models using GRID/CPCA. *J Med Chem* 2001, 44:4072–4081.
118. Afzelius L, Raubacher F, Karlén A, Jorgensen FS, Andersson TB, Masimirembwa CM, Zamora I. Structural analysis of CYP2C9 and CYP2C5 and an evaluation of commonly used molecular modeling techniques. *Drug Metab Disp* 2004, 32:1218–1229.
119. Crivori P, Zamora I, Speed B, Orrenius C, Poggesi I. Model based on GRID-derived descriptors for estimating CYP3A4 enzyme stability of potential drug candidates. *J Comput Aided Mol Des* 2004, 18:155–166.
120. Carosati E. Modelling cytochromes P450 binding modes to predict P450 inhibition, metabolic stability and isoform selectivity. *Drug Discov Today: Technol* 2013 10:e167–e175.
121. MetaSite 3.0, Molecular Discovery Ltd. http://www.moldiscovery.com. Accessed December 2012.
122. Afzelius L, Arnby CH, Broo A, Carlsson L, Isaksson C, Jurva U, Kjellander B, Kolmodin K, Nilsson K, Raubacher F, Weidolf L. State-of-the-art tools for computational site of metabolism predictions: comparative analysis, mechanistical insights, and future applications. *Drug Metab Rev* 2007, 39:61–86.
123. Bonn B, Leandersson C, Fontaine F, Zamora I. Enhanced metabolite identification with MSE and a semi-automated software for structural elucidation. *Rapid Commun Mass Spectrom* 2010, 24:3127–3138.
124. Mass-MetaSite 1.0, Molecular Discovery Ltd. http://www.moldiscovery.com. Accessed December 2012.

