Pharm Res (2013) 30:1458–1463
DOI 10.1007/s11095-013-1012-9

PERSPECTIVE

Emerging Topics in Structure-Based Virtual Screening

Giulio Rastelli

Received: 20 November 2012 / Accepted: 15 February 2013 / Published online: 7 March 2013

ABSTRACT Molecular dynamics simulations and the generation of *ad hoc* chemical libraries are playing an increasingly important and recognized role in structure-based virtual screening. These approaches are important for treating target flexibility and improving the drug discovery pipeline. In this article I will comment on these two topics and put them into perspective.

KEY WORDS ADMET · drug discovery · molecular dynamics · structure-based virtual screening · virtual screening

ABBREVIATIONS

ADMET	adsorption, distribution, metabolism, excretion, toxicity
MM-GBSA	molecular mechanics Generalized Born surface area
MM-PBSA	molecular mechanics Poisson Boltzmann surface area
SBVS	structure-based virtual screening

INTRODUCTION

Designing small molecules with the desired biological activity has long been a desirable but challenging goal. In principle, a deep understanding of the structure and biological function of a target macromolecule, combined with the use of appropriate computational tools able to model the interaction of potential ligands in their binding site, would allow the prediction of biologically active molecules before an experiment is made. In practice, the success of the prediction is often hampered by the complexity of biological macromolecules and by the partial ability of commonly used computational drug design tools to accurately predict active candidates. Therefore, it is not surprising that modeling in drug design is a very active and evolving research field.

Virtual screening is one of the most popular computational approaches for the rapid assessment of large libraries of chemical structures, being able to rationally guide the selection of new biologically active hits. Structure-based virtual screening (SBVS) is usually achieved by using molecular docking tools able to identify complementary orientations of small molecules in the binding site of a target macromolecule and to evaluate the generated docked poses with scoring functions for ligand binding strength prediction (1).

Despite the many approximations, such as the use of approximated scoring functions, the neglect of the entropic term, and the use of rigid receptor structures, we are now in a situation in which SBVS has demonstrated notable potential in discovering biologically active hits (2). Nevertheless, there is pressing need for improving SBVS performance, and for devising new SBVS strategies and approaches able to deliver biologically active compounds with improved potential in both the screen-to-hit and the subsequent hit-to-lead processes. In this perspective article I will discuss two topics that are playing an increasingly important role in SBVS, namely the importance of target flexibility via molecular dynamics simulations, and the important role of assembling *ad hoc* chemical libraries for SBVS. As exemplified in Fig. 1, both topics present important and often interconnected "variables" that need to be considered and properly "tuned" in order to achieve valid starting structures for drug discovery.

THE IMPORTANCE OF TARGET FLEXIBILITY AND THE ROLE OF MOLECULAR DYNAMICS

While in the last 10 years SBVS methods have been improved with respect to the accuracy and efficiency of the

G. Rastelli (✉)
Life Sciences Department, University of Modena and Reggio Emilia
Via Campi 183
41125 Modena, Italy
e-mail: giulio.rastelli@unimore.it

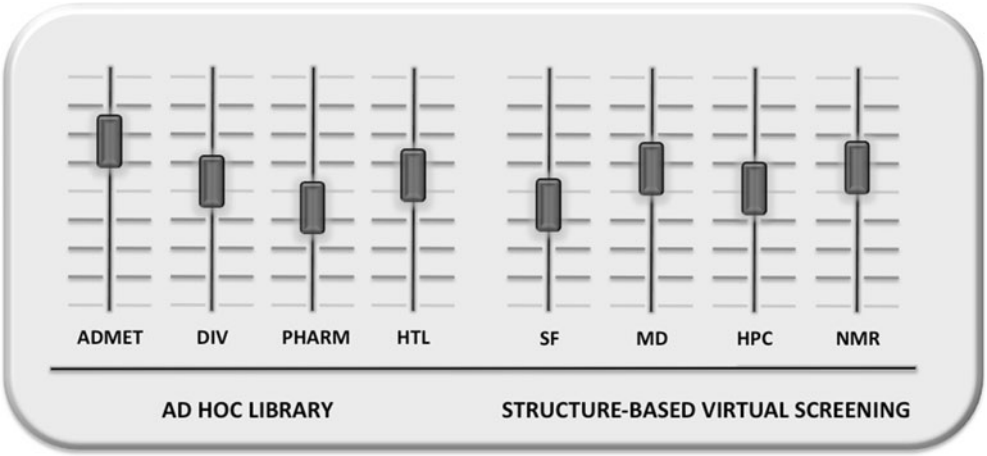

Fig. 1 "Hot features" affecting the quality of SBVS results can be tuned like in a "mixer": ADMET, adsorption, distribution, metabolism, excretion, toxicology; DIV, chemical diversity; PHARM, pharmacophores; HTL, hit-to-lead; SF, scoring functions; MD, molecular dynamics; HPC; high-performance computing; NMR, nuclear magnetic resonance.

available algorithms for pose prediction and ligand scoring, many drawbacks and limitations still exist. For example, scoring functions often fail to estimate ligand binding energies in reasonable agreement with experiment (3–5). Such limitation heavily impacts the selection of molecules for the subsequent biological assays, because the top-fraction of the ranked list turns out to be generally populated by a number of false-positive compounds that are not always easy to detect (6,7). More rigorous free-energy calculations for estimating protein-ligand binding affinities may play a decisive role in overcoming this limitation (8,9). For example, in the last years the use of molecular mechanics Poisson Boltzmann surface area (MM-PBSA) and molecular mechanics Generalized Born surface area (MM-GBSA) methods in rescoring docking solutions has been investigated, and computational schemes have been proposed and applied to test their ability in predicting ligand binding affinities and refining docking poses (10,11 and references therein). These scoring methods have also been integrated with docking and made compatible with automated virtual screening workflows, as for example in the BEAR (Binding Estimation After Refinement) post-docking tool (12). Validation experiments performed on different biological targets showed that, in a number of cases, the MM-PBSA and MM-GBSA scoring functions were able to provide higher enrichment factors with respect to those obtained with other scoring functions implemented in standard high-throughput docking methods (10,11 and references therein).

Besides scoring, one of the most serious challenges in SBVS is how to handle target flexibility. In fact, although docking methods are now able to treat ligand flexibility, most docking methods still consider the protein as a rigid structure, often leading to inaccurate energy evaluation and poor binding mode prediction. This is mainly due to the large number of rotatable bonds (target backbone and side chain bonds) that should be considered in a fully flexible docking approach, which results in high computational costs and limited conformational sampling. Substantial progress in virtual screening will require a deeper understanding of the dynamics of macromolecular complex formation. A practical shortcut that is often used to improve docking calculations is to use multiple fixed receptor conformations, either experimentally determined by crystallography or NMR, or computationally generated (13). Conceptually, docking ligands on multiple protein conformations finds justification from the notion that ligands may select one complementary conformation of the protein out of a pool of pre-existing, different conformations of the protein. In several cases, this approach has led to experimentally validated predictions (14–18, and references therein). However, the use of multiple protein conformations not always provided better results, for example in terms of enrichment of known ligands with respect to molecular decoys (19), indicating that the performances are target-dependent and suggesting that more robust approaches to treat conformational flexibility are probably needed.

One promising possibility to improve conformational sampling in virtual screening is to combine docking with molecular dynamics (MD) simulations. Molecular dynamics has been widely used for investigating macromolecular structures and ligand binding processes. Although MD simulations have inherent approximations, such as the assumption of an additive force field, harmonic terms, the common use of cut-offs in van der Waals and electrostatic interactions, and the use of fixed charges, this technique has proved useful in modeling protein-ligand complexes, refining homology models, performing conformational analyses, highlighting time-dependent structural features relevant for binding, and predicting protein conformational changes (20,21). Modeling of dynamic molecular features relevant

for binding, not currently achievable with current SBVS methods, could improve the accuracy of predictions of novel bioactive compounds, and may give useful insights into the molecular interactions that the ligands establish with their target protein before further experiments are made. Conformational sampling of protein-ligand complexes with MD may reveal new binding modes, predict new protein conformations, and may highlight allosteric effects that in turn affect target activity and protein-ligand interactions. Moreover, they can highlight cryptic binding sites, that are not immediately obvious from the analysis of crystal structures, but that can turn out to be druggable and therefore useful for drug discovery.

Recently, a relaxed complex scheme (RCS) approach in which potential ligands are docked into multiple protein conformations extracted from a MD trajectory has been proposed (22). In this approach, each potential ligand is associated with a whole spectrum of energy scores, one for each protein conformation used for docking. In other approaches, MD simulations have been started from one or more docked complexes and then run for a certain simulation time, using various computational protocols. The resulting MD trajectories have been used *i*) to evaluate ligand-induced conformational changes, *ii*) to assess whether the docked orientation is reasonably stable over time, and *iii*) to perform binding free energy predictions (10,11,23). Conceptually, this last approach implies that ligands may not only be able to select pre-existing conformational states of the protein, as pointed out above, but also to induce ligand-dependent conformational changes as a result of conformational induced-fit effects arising from the molecular dynamics of the complex. A major issue of this approach is how to make the (generally high) computational costs of MD simulations compatible with the dimensions of a typical SBVS, in which libraries containing a huge number of compounds are usually screened. We recently proposed that the refinement of docked complexes by means of an energy minimization and a short molecular dynamics simulation, followed by MM-PBSA and/or MM-GBSA binding free energy predictions on the refined complexes, is a useful strategy to post-process docking solutions (for a review, see (10)). This approach is the one implemented in the automated workflow BEAR (12). It should be emphasized that biological processes occur on generally long time scales, and an accurate conformational sampling of protein-ligand complexes would require extended, and sometimes impractical, simulation times. A number of possible solutions to this issue, based on 'enhanced sampling' methods, have been proposed (24). These include replica exchange MD (REMD), accelerated MD, umbrella sampling, and coarse-grained MD. In addition, parallel and distributed computing is providing unprecedented resources to run MD simulations the length that was not affordable before. Noteworthy, the advent of GPUs is making it possible to run longer explicit solvent MD simulations in exceedingly short times. Therefore, MD simulations are likely to play an increasingly important role in SBVS and drug discovery in general. Future implementations of enhanced sampling MD methods in SBVS workflows will be of particular interest, and these in turn will take advantage of the power of high-performance computing systems like GPUs. Likewise, further improvements in force fields, in particular in the development of polarizable force fields (25), will likely lead to a better description of the dynamical aspects of protein-ligand structures and therefore to an improved agreement with experimental binding affinities. Recently, considerable attention has been given to the comparison between experiment and simulation for macromolecules, allowing a more thorough force field validation and, ultimately, a higher accuracy in modeling protein structures and protein complexes (26,27). In particular, while comparisons between MD simulations and solution state NMR experiments has been especially useful in force field validation (28,29), the use of NMR data may also be prospected for testing and improving virtual screening methods (30). It is common practice to test the accuracy of a docking protocol by performing re-docking experiments of a co-crystallized ligand, using the RMSD to a crystal binding pose as the sole measure of docking quality. If these validations were extended to NMR data, this might then highlight multiple, sometimes unexpected, binding modes. Such analyses might therefore help capturing the dynamic aspects of the protein-ligand interaction.

THE IMPORTANCE OF *AD HOC* CHEMICAL LIBRARIES IN SBVS

Several chemical libraries and public domain chemistry databases for SBVS investigations are available (31,32). One of the most exploited strategies in SBVS is the use of collections of commercially available compounds (33). Once the best hits are selected, these can be forwarded directly to biological testing without the need of chemical synthesis. Alternatively, chemical libraries for SBVS can be prepared by assembling collections of virtual compounds based on validated synthetic protocols and available starting materials. These libraries are encoded using a list of reactants that is a superset of that used in the actual synthesis, therefore greatly increasing the chemical space covered by the virtual compounds compared to the synthesized compounds (34,35). Because the virtual hits are annotated with a validated synthetic protocol, the chemistry turnaround time is expected to be shorter. This allows researchers to drive the SBVS process toward compounds that can be readily synthesized, and for which a rapid exploration of structure-activity relationships is possible.

Chemical libraries typically used in SBVS contain from thousands to million compounds. Of particular importance is the preparation of the chemical library. For each compound, different tautomers, enantiomers and possible ionization states at physiological pH should be included. Moreover, a conformational analysis may be conducted on all structures, and the database may include conformers within a certain energy threshold from the global energy minimum. It is important to note that all the operations described above increase significantly the number of structures included in the database, making the subsequent SBVS process more computationally demanding. In these cases, filtering the database prior to SBVS with *ad hoc* pharmacophores may be advisable in order to rationally reduce the number of compounds to be processed in the following steps (36). In this context, building pharmacophore models from three-dimensional structures of the target protein(s) is a valuable approach to account for key elements of protein-ligand binding interactions (37–39). By explicitly requesting structural features or molecular properties deemed important for binding, such an approach is able to remove compounds that do not possess the requested features. Pharmacophores are problem-dependent and should be built on the basis of specific requirements (40). When the objective is the identification of new and chemically diverse biologically active hits from large libraries of compounds, it may be advisable to screen with pharmacophores encoding few but key and conserved features absolutely required for binding. Then, the resulting filtered compounds are forwarded to SBVS workflows for further screening and narrower selection (41). Conversely, if the aim is to find compounds with a desired selectivity profile, more specific features (for example interactions with less conserved portions of the binding site, excluded volumes, or desired nature of intermolecular interactions) should be included (39,42). In both cases, it may be useful to include ligand-specific features encoded into more than one crystal structure.

Evaluating drug-like properties of compounds in a chemical library is another crucial aspect. Given that biological activity and ADMET properties often tend to have diametrically opposed relationship with physicochemical properties, optimizing *in vitro* potency and ADMET properties is usually a daunting task (43). For this reason, it is becoming increasingly recognized that selecting hits with lower molecular weight, even with lower *in vitro* potency (for example micromolar instead of nanomolar or subnanomolar) but with favorable drug-like properties, might give more chances of developing an orally administered drug. Given the considerable difficulty of optimizing ADMET properties in drug discovery, physicochemical properties of the compounds in the chemical libraries should be carefully evaluated and considered early in the SBVS process.

When building a chemical library for SBVS, compounds with poor drug-like properties should be filtered out, in order to exclude *a priori* compounds with clearly unfavorable pharmacokinetic properties. Moreover, properties that could interfere with the subsequent *in vitro* and *in vivo* experimentation, like compound solubility, the presence of highly reactive groups or groups that appear as frequent hitters (promiscuous compounds) in many biochemical high-throughput screens should be taken into account. In this way, the risk of selecting compounds with poor absorption and permeation, or compounds whose biological inactivity could be attributed to insolubility in the assay medium or low membrane permeation, as well as compounds whose biological activity derives from unspecific mechanisms, may be reduced. Several drug- and lead-likeness simple rules are available for this purpose (44). Moreover, a number of studies have examined the relationship between computed molecular properties and *in vivo* and *in vitro* ADMET properties (45). *In silico* modeling of ADMET properties will play an increasingly important role in chemical library design and screening (46). Very recently, it was suggested that exploiting pharmacological similarity can provide useful insight into potential safety liabilities that are neither identified by pure chemical similarity nor by risk assessment on individual targets (47). The Biosim web server has thus been developed to aid in the prediction of specific off-targets and to highlight potential safety liabilities via pharmacological similarity concepts (47). It is expected that further improvements of the performance of ADMET predictions will generate better predictive models and, ultimately, more promising compounds to be passed to *in silico* screening. Importantly, strategies specifically devised to better integrate ADMET predictions and SBVS methods may synergistically improve the quality of the generated hits and make the subsequent drug optimization process less difficult.

Another problem, often underestimated, in drug development is that a candidate drug should have negligible (or at least controlled) off-target effects. *In silico* approaches have been put forward to predict off-target and/or multi-target activities as well as drug repositioning (15,48,49). In line with this direction, the public-private PHACTS consortium (www.openphacts.org) is building an Open Pharmacological Space for providing a comprehensive framework of chemical, biological and pharmacological information on drugs and small molecules for screening. This semantically enriched and fully interoperable platform will deliver information on small molecules and their pharmacological profiles, including pharmacokinetics and ADMET data as well as on biological targets and pathways. When the platform will be made available, it may provide valuable information for generating focused libraries of *ad hoc* compounds based on these concepts to be used in virtual screening. One final comment is worth for the chemical diversity of the library, an important feature that influences SBVS results. If the aim is to find a number of chemically diverse hits, it is generally

advisable to remove from the library compounds with similar structures, in order to avoid populating the top positions of the ranked lists with compounds carrying a similar structural information content (*e.g.* compounds typically differing for substituents or isosteric replacements) (50). This makes the selection of chemically diverse candidates less difficult. However, it should be noted that when SBVS methods are used for hit expansion or analog searching, chemical diversity should be limited and generally restricted to the desired classes of compounds (51).

Although it has become generally easy to find biologically active compounds (although usually not ultra potent) for a given macromolecular target using SBVS techniques, it should be emphasized that the optimization of the resulting hits in the subsequent hit-to-lead process remains generally difficult (51). This poses a major problem in academia and pharmaceutical companies, because a large number of potentially interesting hit candidates are discovered but then remain largely unexplored. Major obstacles that hamper compound's follow up include difficult synthetic schemes and undesirable ADMET properties. As for synthetic difficulties, virtual synthesis and retrosynthetic schemes should be evaluated already during the analysis of the SBVS results, in order to select, among the pool of potential ligands present in the ranked lists, those that are more amenable to optimization. A better convergence of virtual screening (and modeling in general) with practical experimental tools and techniques may provide a more sustained success. Approaches such as diversity-oriented synthesis, click-type assembly of ligands into the protein binding site, and biologically-oriented synthesis, are examples in this direction (52,53). One final obstacle to the prosecution of the hit-to-lead optimization process, which is especially recognized by pharma companies, is that the identified hits may not possess enough structural novelty to warrant patentability. This is one case in which a careful design of the chemical library used for *in silico* screening may play a decisive role, for example by assembling compounds from fragments using chemical diversity and novelty criteria.

REFERENCES

1. Kitchen DB, Decornez H, Furr JR, Bajorath J. Docking and scoring in virtual screening for drug discovery: methods and applications. Nat Rev Drug Discov. 2004;3(11):935–49.
2. Ripphausen P, Nisius B, Peltason L, Bajorath J. Quo vadis, virtual screening? A comprehensive survey of prospective applications. J Med Chem. 2010;53(24):8461–7.
3. Cheng T, Li X, Li Y, Liu Z, Wang R. Comparative assessment of scoring functions on a diverse test set. J Chem Inform Model. 2009;49(4):1079–93.
4. Waszkowycz B. Towards improving compound selection in structure-based virtual screening. Drug Disc Today. 2008;13(5–6):219–26.
5. Plewczynski D, Łaźniewski M, Augustyniak R, Ginalski K. Can we trust docking results? Evaluation of seven commonly used programs on PDBbind database. J Comput Chem. 2011;32(4):742–55.
6. Šink R, Gobec S, Pečar S, Zega A. False positives in the early stages of drug discovery. Curr Med Chem. 2010;17(34):4231–55.
7. Coan KED, Shoichet BK. Stoichiometry and physical chemistry of promiscuous aggregate-based inhibitors. J Am Chem Soc. 2008;130(29):9606–12.
8. Michel J, Essex JW. Prediction of protein-ligand binding affinity by free energy simulations: assumptions, pitfalls and expectations. J Comput-Aid Mol Des. 2010;24(8):639–58.
9. Jorgensen WL Efficient drug lead discovery and optimization. Acc Chem Res. 2009; 42(6):724-733
10. Parenti MD, Rastelli G. Advances and applications of binding affinity prediction methods in drug discovery. Biotechnol Adv. 2012;30(1):244–50.
11. Rastelli G, Del Rio A, Degliesposti G, Sgobba M. Fast and accurate predictions of relative binding free energies using MM-PBSA and MM-GBSA. J Comp Chem. 2010;31(4):797–810.
12. Degliesposti G, Portioli C, Parenti MD, Rastelli G. BEAR, a novel virtual screening methodology for drug discovery. J Biomol Screen. 2011;16(1):129–33.
13. Totrov M, Abagyan R. Flexible ligand docking to multiple receptor conformations: a practical alternative. Curr Opin Struct Biol. 2008;18(2):178–84.
14. Ferrari AM, Wei BQ, Costantino L, Shoichet BK. Soft docking and multiple receptor conformations in virtual screening. J Med Chem. 2004;47(21):5076–84.
15. Kirchmair J, Distinto S, Schuster D, Spitzer G, Langer T, Wolber G. Enhancing drug discovery through *in silico* screening: strategies to increase true positives retrieval rates. Curr Med Chem. 2008;15(20):2040–53.
16. Polgár T, Baki A, Szendrei GI, Keseru GM. Comparative virtual and experimental high-throughput screening for glycogen synthase kinase-3beta inhibitors. J Med Chem. 2005;48(25):7946–59.
17. Rao S, Sanschagrin PC, Greenwood JR, Repasky MP, Sherman W. Farid R Improving database enrichment through ensemble docking. J Comput Aided Mol Des. 2008;22(9):621–7.
18. Bowman AL, Nikolovska-Coleska Z, Zhong H, Wang S, Carlson HA. Small molecule inhibitors of the MDM2-p53 interaction discovered by ensemble-based receptor models. J Am Chem Soc. 2007;129(42):12809–14.
19. Sgobba M, Caporuscio F, Anighoro A, Portioli C, Rastelli G. Application of a post-docking procedure based on MM-PBSA and MM-GBSA on single and multiple protein conformations. Eur J Med Chem. 2012;58:431-440.
20. Wang W, Donini O, Reyes CM, Kollman PA. Biomolecular simulations: Recent developments in force fields, simulations of enzyme catalysis, protein-ligand, protein-protein, and protein-nucleic acid noncovalent interactions. Annu Rev Biophys Biomol Struct. 2001;30:211-243.
21. Karplus M, Mc Cammon JA. Molecular dynamics simulations of biomolecules. Nat Struct Biol. 2002;9(9):646–52.
22. Amaro RE, Baron R, McCammon JA. An improved relaxed complex scheme for receptor flerxibility in computer-aided drug design. J Comput Aided Mol Des. 2008;22(9):693–705.
23. Homeyer N, Gohlke H. Free energy calculations by the molecular mechanics Poisson-Boltzmann surface area method. Mol Inf. 2012;31:114–22.
24. Liwo A, Czaplewski C, Ołdziej S, Scheraga HA. Computational techniques for efficient conformational sampling of proteins. Curr Opin Struct Biol. 2008;18:134–9.
25. Wang J, Cieplak P, Li J, Wang J, Cai Q, Hsieh M, *et al.* Development of polarizable models for molecular mechanical calculations II: Induced dipole models significantly improve

accuracy of intermolecular interaction energies. J Phys Chem B. 2001;115(12):3100–11.
26. Schaeffer RD, Fersht A, Daggett V. Combining experiment and simulation in protein folding: closing the gap for small model systems. Curr Opin Struct Biol. 2008;18(1):4–9.
27. Putnam CD, Hammel M, Hura GL, Tainer JA. X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. Q Rev Biophys. 2007;40(3):191–285.
28. Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, Simmerling C. Comparison of multiple Amber force fields and development of improved protein backbone parameters. Proteins. 2006;65(3):712–25.
29. Klepeis JL, Lindorff-Larsen K, Odror R, Shaw DE. Long-timescale molecular dynamics simulations of protein structure and function. Curr Opin Struct Biol. 2009;19:120–7.
30. Damm KL, Carlson HA. Exploring experimental sources of multiple protein conformations in structure-based drug design. J Am Chem Soc. 2007;129(26):8225–35.
31. Williams AJ, Ekins S, Tkachenko V. Towards a gold standard: regarding quality in public domain chemistry databases and approaches to improving the situation. Drug Discov Today. 2012;17(13/14):685–701.
32. Nicola G, Liu T, Gilson MK. Public domain databases for medicinal chemistry. J Med Chem. 2012;55(16):6987–7002.
33. Irwin JJ, Shoichet BK. ZINC – A free database of commercially available compounds for virtual screening. J Chem Inf Model. 2005;45(1):177–82.
34. Vainio MJ, Kogej T, Raubacher F. Automated recycling of chemistry for virtual screening and library design. J Chem Inf Model. 2012;52:1777–86.
35. Hu Q, Peng Z, Kostrowicki J, Kubi A. LEAP into the Pfizer global virtual library (PGVL) space: creation of readily synthesizable design ideas automatically. In: Zhou JZ, editor. Chemical library design. Methods in molecular biology, vol. 685. New York: Humana Press; 2011. p. 253–76.
36. Caporuscio F, Tafi A. Pharmacophore modeling: a forty year old approach and its modern synergies. Curr Med Chem. 2011;18(17):2543–53.
37. Heinke R, Spannhoff A, Meier R, Trojer P, Bauer I, Jung M, *et al.* Virtual screening and biological characterization of novel histone arginine methyltransferase PRMT1 inhibitors. ChemMedChem. 2009;4(1):69–77.
38. Knox AJS, Price T, Pawlak M, Golfis G, Flood CT, Fayne D, *et al.* Integration of ligand and structure-based virtual screening for the identification of the first dual targeting agent for heat shock protein 90 (Hsp90) and tubulin. J Med Chem. 2009;52(8):2177–80.
39. Rastelli G, Pacchioni S, Sirawaraporn W, Sirawaraporn R, Parenti MD, Ferrari AM. Docking and database screening reveal new classes of Plasmodium falciparum dihydrofolate reductase inhibitors. J Med Chem. 2003;46(14):2834–45.
40. Langer T. Pharmacophores in drug research. Mol Inform. 2010;29(6–7):470–5.
41. Shangary S, Wang S. Small-molecule inhibitors of the MDM2-p53 protein-protein interaction to reactivate p53 function: a novel approach for cancer therapy. Annu Rev Pharmacol Toxicol. 2009;49:223–41.
42. Mazanetz MP, Fischer PM. Untangling tau hyperphosphorylation in drug design for neurodegenerative diseases. Nat Rev Drug Discov. 2007;6(6):464–79.
43. Gleeson MP, Hersey A, Montanari D, Overington J. Probing the links between *in vitro* potency, ADMET and physicochemical parameters. Nat Rev Drug Discov. 2010;10(3):197–208.
44. Chuprina A, Lukin O, Demoiseaux R, Buzko A, Shivanyuk A. Drug- and lead-likeness, target class, and molecular diversity analysis of 7.9 million commercially available organic compounds provided by 29 suppliers. J Chem Inf Model. 2010;50(4):470–9.
45. Sutherland JJ, Raymond JW, Stevens JL, Baker TK, Watson DE. Relating molecular properties and *in vitro* assay results to *in vivo* drug disposition and toxicity outcomes. J Med Chem. 2012;55(4):6455–66.
46. Kortagere S, Ekins S. Troubleshooting computational methods in drug discovery. J Pharmacol Toxicol Methods. 2010;61(2):67–75.
47. Muthas D, Boyer S. Exploiting pharmacological similarity to identify safety concerns – listen to what the data tells you. Mol Inform. 2013;32(1):37–45.
48. Taboureau O, Baell JB, Fernández-Recio J, Villoutreix BO. Established and emerging trends in computational drug discovery in the structural genomics era. Chem Biol. 2012;19(1):29–41.
49. Keiser MJ, Setola V, Irwin JJ, Laggner C, Abbas AI, Hufeisen SJ, *et al.* Predicting new molecular targets for known drugs. Nature. 2009;462(7270):175–81.
50. Medina-Franco JL, Martínez-Mayorga K, Giulianotti MA, Houghten RA, Pinilla C. Visualization of the chemical space in drug discovery. Curr Comput-Aided Drug Des. 2008;4(4):322–33.
51. Manly CJ, Chandrasekhar J, Ochterski JW, Hammer JD, Warfield BB. Strategies and tactics for optimizing the Hit-to-Lead process and beyond–a computational chemistry perspective. Drug Discov Today. 2008;13(3–4):99–109.
52. Nielsen TE, Schreiber SL. Towards the optimal screening collection: a synthesis strategy. Angew Chem Int Ed. 2008;47(1):48–56.
53. Dandapani S, Marcaurelle LA. Current strategies for diversity-oriented synthesis. Curr Opin Chem Biol. 2010;14(3):362–70.