



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



Large-scale integrated super-computing platform for next generation virtual drug discovery

Wayne Mitchell^{1,2} and Shunji Matsumoto³

Traditional drug discovery starts by experimentally screening chemical libraries to find hit compounds that bind to protein targets, modulating their activity. Subsequent rounds of iterative chemical derivitization and rescreening are conducted to enhance the potency, selectivity, and pharmacological properties of hit compounds. Although computational docking of ligands to targets has been used to augment the empirical discovery process, its historical effectiveness has been limited because of the poor correlation of ligand dock scores and experimentally determined binding constants. Recent progress in super-computing, coupled to theoretical insights, allows the calculation of the Gibbs free energy, and therefore accurate binding constants, for usually large ligand–receptor systems. This advance extends the potential of virtual drug discovery. A specific embodiment of the technology, integrating *de novo*, abstract fragment based drug design, sophisticated molecular simulation, and the ability to calculate thermodynamic binding constants with unprecedented accuracy, are discussed.

Addresses

¹ Experimental Therapeutics Centre, Agency for Science and Technology Research (A*STAR), 31 Biopolis Way, #03-01 Nanos, Biopolis, Singapore 138669, Singapore

² Nanyang Technological University, School of Computer Engineering, Division of Information Sciences, 50 Nanyang Avenue, Singapore 639798, Singapore

³ *In-silico* Drug Discovery Research Division, Bio-IT Business Development Unit, Fujitsu Limited, 1-9-8 Nakase, Mihama-ku, Chiba-city, Chiba 261-8588, Japan

Corresponding author: Matsumoto, Shunji (shun@jp.fujitsu.com)

Current Opinion in Chemical Biology 2011, 15:553–559

This review comes from a themed issue on
Next Generation Therapeutics
Edited by Alex Matter and Thomas H. Keller

Available online 30th June 2011

1367-5931/\$ – see front matter
© 2011 Elsevier Ltd. All rights reserved.

DOI 10.1016/j.cbpa.2011.06.005

Introduction

Drug discovery is distressed. The number of approved new molecular entities has declined steadily for 15 years [1], the cost per new approved compound has breached the one billion United States Dollar (USD) benchmark [2] and, by one informed estimate, the financial return provided by all therapeutic product categories does not even recover the capital costs of their development.

Moreover, when biologics are removed from this model, the net present value of pharmaceutical Research and Development (R&D) investment is actually negative [3]. Disturbingly, poor research performance occurs in the context of increasing social need. The rising incidence of drug resistant bacteria is well documented [4], viral pathogens such as Severe Acute Respiratory Syndrome (SARS), Dengue and H1N1 pose pandemic threats [5], and an aging and more affluent global population drives up the prevalence of chronic diseases that are if anything more difficult than infectious diseases to drug. In the United States, where by 2023 the Census Bureau projects ~15% of males and ~19% of females will be 65 years old or older, rates of cancer are expected to rise from 33% to 62%, of diabetes from 33% to 53%, of cardiovascular complaints from 6% to 39%, of mental disorders from 35% to 54%, for a total increase in chronic morbidity in the U.S. population from 17% to 42% [6]. An increase in successful drug discovery, especially against difficult chronic targets, is clearly desirable. However, progress must materialize in the framework of reduced per compound cost and enhanced efficiency, since capital will not continue to flow into a sector that offers net negative returns. Computational or ‘virtual’ drug discovery strategies, potentially cheaper and faster, offer attractive alternative, or at least complimentary, routes to improved R&D performance in the therapeutics sector [7].

Computational drug discovery

The physiological effect of a drug is mediated by electrostatic and geometrical interactions of the atoms of the ligand with the atoms of its corresponding receptor, interactions which conform to the laws of physics and quantum chemistry, and which can therefore be described by predictive mathematical models [8^{*}]. Although these models are complex (and in the quantum case inherently non-exact), researchers active in the computer intensive field of molecular graphics realized thirty years ago that *in silico* assessment of drug-receptor binding could be deployed to accelerate drug discovery [9,10]. They also realized that continuing operation of Moore’s Law, with 18 month doublings in computing efficiency and economy, implied an on-going improvement and refinement in computational techniques. The crux of the computational drug discovery paradigm is this coupling of the fundamental laws of biophysics with the accelerating technological performance of the semiconductor industry. The former inspires faith that the approach can work in theory, the latter that it will work in practice.

Virtual compound screening

Virtual screening, whether of compounds or molecular fragments, has two stages. First, the algorithms attempt to find the correct conformation and position the ligand in the active site of the receptor, and then they try to quantify the quality of particular atomic arrangements by assigning a score. Several technically different approaches to predicting ligand–receptor interactions have been developed, but all are known as ‘docking’ algorithms after the suggestively named primogenitor program, ‘DOCK’ [9]. The modeling of ligand–receptor atomic interactions presupposes an accurate three-dimensional molecular structure of the receptor so that inter-atomic forces can be calculated. Since protein folding cannot yet be modeled, this means having an X-ray crystal or NMR structure of the receptor, or a homology model which maps a related protein sequence onto a known structure. *A priori* one might suppose the experimentally determined crystal structure to be inevitably superior. Surprisingly, a meta-analysis of DOCKing studies concluded that in some cases virtual screening was more successful on homology models compared to experimental structures [11**]. This seems counter-intuitive, but may indicate that the relaxed precision of the homology models indirectly capture conformational flexibility that is lost in ‘frozen’, possibly subtly distorted crystal structures. In any case, an homology model must start from a closely related experimental structure, so an important contributing factor in the increased utility of computational drug discovery is the rapid growth in the number of available protein structures (currently approaching 75,000 structures in the Protein Data Base <http://www.pdb.org/pdb/statistics/contentGrowthChart.do?content=total&seqid=100>), a number which in turn reflects improvements in protein production, robotic crystallization regimens, and the wide availability of sophisticated advanced light sources [12]. Another positive development in virtual screening infrastructure is the creation of curated virtual compound databases that provide large prebuilt sets of virtual representations of commercially available molecules suitable for input to virtual screens. ZINC at the University of California San Francisco [13], and EDU-LISS at Edinburgh University [14], are two examples. The much larger Chemical Universe Database GDB-13 takes a different approach, attempting to construct the universe of ‘synthetically plausible’, rather than ‘available’ compounds [15].

Molecular docking: successes and limitations

At a high level the performance of dock programs can be measured by two criteria: ‘DOCKing power’ (the ability to identify the correct experimental ligand binding pose in a collection of incorrect, computer generated ‘decoy’ poses, i.e. the ability to correctly position ligands in the active site, or to ‘pose’ them); and ‘scoring power’ (the ability to produce dock binding scores that correlate with

experimentally determined binding affinities). In the past decade a large number of comparative studies of the performance of various dock programs have been undertaken, in both academic and pharmaceutical settings [11,16,17**,18*,19–22]. Despite the diverse backgrounds of the investigators, and although these studies differ in methodology and are not directly comparable, they nevertheless unanimously agree on two points. One, dock algorithms fairly accurately *pose* ligands in the active site, and two, the same dock algorithms poorly *score* those ligands’ affinity. In other words, dock programs correctly identify the geometry of ligand–receptor systems, but, do not in general accurately predict the binding energy, and therefore cannot predict ligand potency. To make this concrete, a typical dock screen might produce 1000 ‘hit’ compounds, but, the most potent compounds are as likely to be ranked at the bottom of that list by the scoring function as they are to appear near the top. This is a significant deficiency since the expected potency of a compound often will be the operational feature of interest, for example in prioritizing compounds for medicinal chemistry. In sum, Dock algorithms can ‘pose’ ligands well but they ‘score’ them poorly.

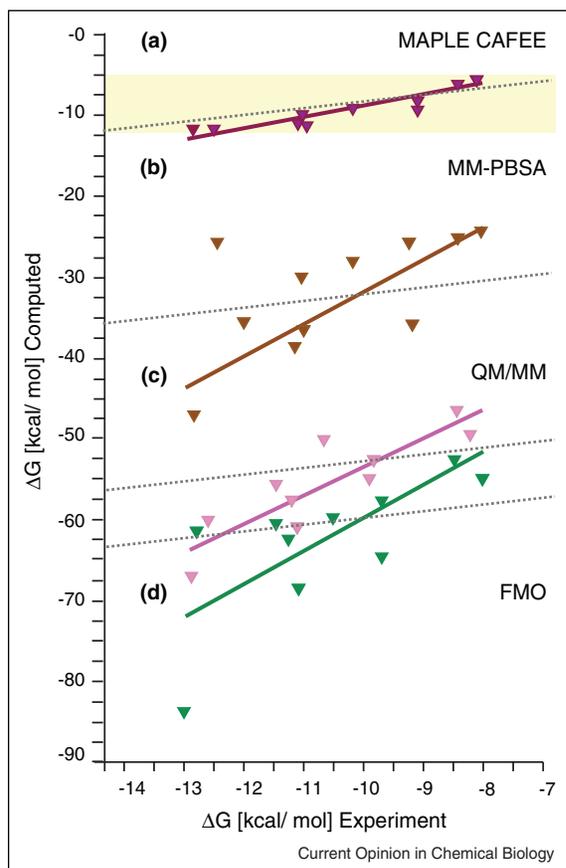
Beyond dock scores: accurate binding affinity from thermodynamic calculation with MAPLE CAFEE

$\Delta G = RT \ln K_d$ exactly relates the computed Gibbs free energy difference ΔG and experimentally measured dissociative constant K_d under temperature T (where, R is the gas constant). Free energy differences between bound and unbound equilibrium states of a protein–ligand–water system (ΔG) gives the binding affinity of the ligand, which in general translates into drug efficacy. In other words, *correct* computation of ΔG values for a series of ligands leads immediately to a ligand list accurately ranked by potency.

Computational methods to perform the ΔG calculation have been studied enthusiastically since the late 1990s when it was proved that a nonequilibrium process in finite-time can derive the binding free energy exactly [23–25]. This theoretical insight was followed up in 2005 when ΔG was shown to be approachable by massively parallel computation. Scaling up to thousands of concurrent CPUs reduced the computational requirement from years to days, and allowed binding free energy of real molecular systems to be computed. Access to this high performance computational resource made possible an important series of proof of concept experiments that in turn produced calculated binding affinities in excellent agreement with corresponding experimental values [26,27]. Subsequently, the computational methodology has been improved, a better force field refining method has been implemented, and the platform, christened Massively Parallel Computation of Absolute binding Free Energy with well Equilibrated system (MAPLE

CAFEE), has been further validated in follow-on projects [28–30]. Useful thermodynamic calculation comes with a heavy computational burden (an example of the trade-offs between speed and accuracy that turn up in many computational problems), but recent advances in the construction of large-scale computing environments have shrunk the envelope of computational time and brought these methods into the realm of practical application ([31], <http://www.nsc.riken.jp/index-eng.html>, <http://www.fujitsu.com/global/news/pr/archives/month/2010/20100928-01.html>). Cumulatively, improved computational methodology, enhanced infrastructure, and theoretical advances have combined to achieve chemical thermodynamic calculations that are sufficiently accurate to provide reliably ordered and prioritized lists of hit compounds, either from virtual libraries or *de novo* design.

Figure 1



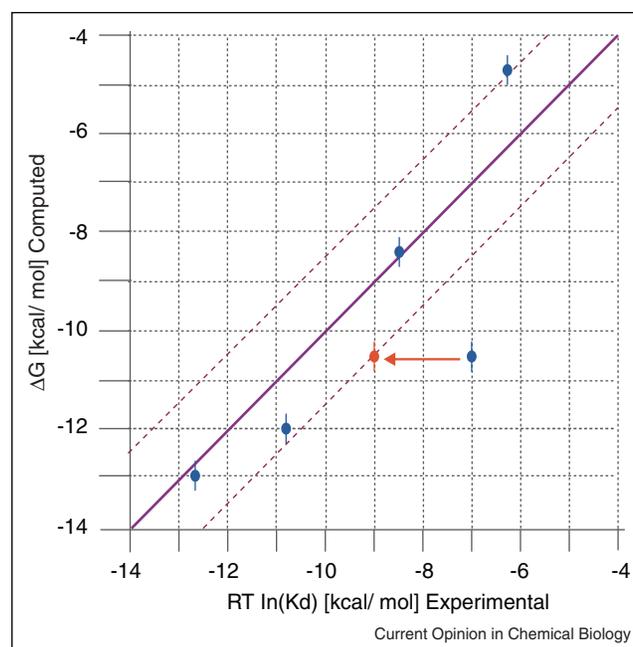
Controlled comparison of four binding affinity prediction methods on the same 17,000 atom FK506-FKBP data set. Each graph shows the correlation between observed (experimental) binding affinity (on the horizontal axis) and computed binding free energy (on the vertical axis). The dotted lines show a slope of 1 (perfect correlation, with possible offset). The yellow box shows the area with no X–Y offset. **(a)** Massively Parallel Computation of Absolute binding Free Energy with Well Equilibrated system (MAPLE CAFEE). **(b)** Molecular Mechanics Poisson–Boltzmann Surface Area (MM-PBSA). **(c)** Quantum Mechanics/Molecular Mechanics (QM/MM). **(d)** Fragment Molecular Orbital Method (FMO).

Comparison of binding affinity calculation by four methods

In 2008, three separate organizations compared binding affinity prediction methods on a defined data set using identical force field parameters [32]. The system of interest was FK Binding Protein (FKBP)–ligand–water of about 17,000 atoms, with water explicitly modeled with TIP3P. A co-crystal structure of the FKBP-FK506 complex was obtained from the Protein Data Bank (PDB). Then other ligands replaced FK506 one by one. The four methods tested were Fragment Molecular Orbital method (FMO), Molecular Mechanics Poisson–Boltzmann Surface Area (MM-PBSA), the hybrid Quantum Mechanics/Molecular Mechanics (QM/MM) and MAPLE CAFEE.

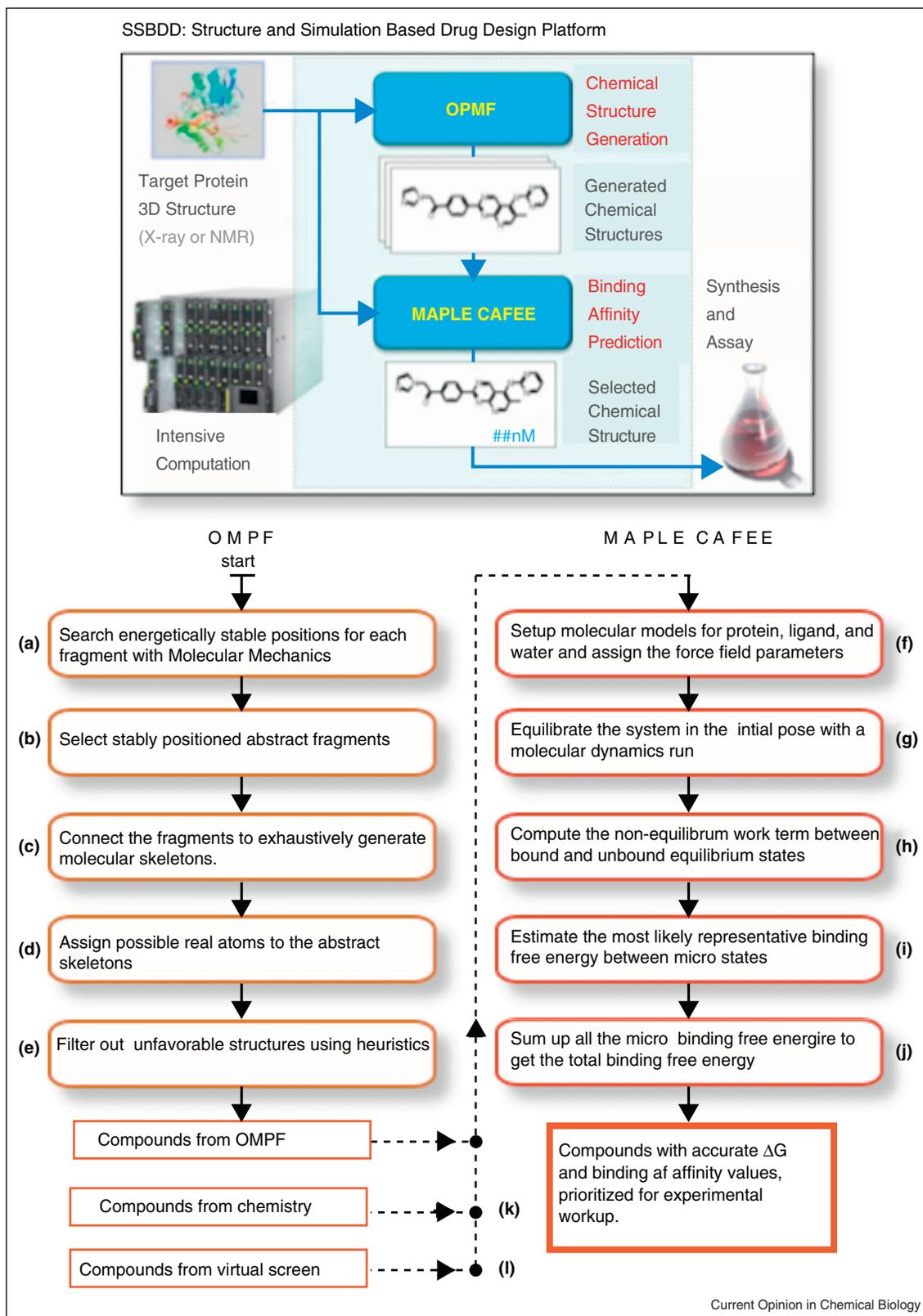
Figure 1a–d, reprinted from [32], shows the correlation between observed binding affinity and computed binding free energy for the four methods. While the computational time required for each method conformed to expectation, and each computational method produced results that correlate somewhat with experimental values, there were significant differences in the scale and shifts from the origin of the vertical axis. MAPLE CAFEE produced the best agreement with a difference from observed affinities within 0.5 [kcal/mol] (an error of 1.4 [kcal/mol] is a good benchmark for computational accuracy since the error of experimentally measured affinity is in this range).

Figure 2



Blind-test of MAPLE CAFEE. The binding energies of five ligands on a cancer related protein were calculated by MAPLE CAFEE, then plotted against the experimentally determined binding energies. The dotted lines define a band within 1.5 kcal of perfect correlation. The binding constant of the outlier was measured a second time in a repeat experiment. The recalculated value is shown by the red arrow.

Figure 3



The OPMF and MAPLE CAFEE workflows. The standard drug-design process of OPMF comprises five steps. **(a)** Search energetically stable positions for each abstract fragment in/on a targeted protein with molecular mechanics simulation (MM). At the discretion of the investigator, physically discovered molecular fragments, for example from NMR or crystal soaks, can be substituted for the virtual fragment inputs. **(b)** Select a stably positioned set of

A blind challenge of MAPLE CAFEE

In 2009, MAPLE CAFEE was challenged by a pharmaceutical industry collaborator to a blind-test of binding affinity prediction (unpublished results). The algorithm was given the structure of a cancer related protein and five ligands, one a co-crystal structure, three without co-crystal data but where the small molecule was similar to the first, and a final ligand without co-crystal data and structurally dissimilar to the others. The experimentally determined binding affinities were hidden. As seen in Figure 2 the results showed very good agreement between binding free energies computed by MAPLE CAFEE, and experimentally measured dissociation constants. There was one outlier, however, the pharmaceutical partner suggested this particular experimental measurement was unreliable, and indeed, when repeated, the experimental value fell within 1.4 kcal/mol of the computed value.

A coherent structure based drug-design platform

MAPLE CAFEE is part of a larger Structure and Simulation Based Drug-Design (SSBDD) platform (Figure 3), that also includes the Optimum Packing of Molecular Fragments (OPMF) module, an abstract fragment based, *de novo*, drug-design tool (Figure 3a–e). Each component of SSBDD can be applied to specific purposes independently, but the platform is designed as an integrated system for creating novel and active chemical entities.

OPMF is a virtual fragment based *de novo* drug-design tool that generates chemical structures predicted to bind and modulate the activity of target proteins whose 3D structures are known. In order to examine the vast virtual chemical space efficiently, OPMF generalizes real-atom fragments to abstract conceptual fragments at the first step. This strategy allows a small number of abstract fragments to represent thousands of real fragments. For example, benzene, pyridine, pyrimidine, triazine are generalized to an abstract six-membered aromatic ring built of abstract atoms having representative diameter and properties. The approximation is reversed later when abstract structures are replaced by precisely assigning real-atoms that rigorously capture exact physical features including electrostatics, orbitals and van der Waals radii.

MAPLE CAFEE is a binding affinity prediction system, as already mentioned, that computes ΔG as precisely as possible using a massively parallel sampling method reinforced by many molecular dynamic (MD) simulations [26] (Figure 2f–j). Parallelization, via the Message Passing Interface (MPI) plays a critical role in the system equilibration step of MAPLE CAFEE (Figure 3g), for example by allowing information such as hydrogen bond frequency to be derived and fed back into the drug-design process. Computing the nonequilibrium work term (Figure 3h) invokes numerous MD jobs, and this is made practical by adopting a massively parallel approach. Each MD job generally takes a couple of days on a single CPU core, but MPI parallelism for a microstate MD job would reduce the computational time almost in proportion to the number of cores. Therefore, if 10,000 CPU cores were available, it would be possible to examine hundreds of novel structures in a week without any actual synthesis. This is the biggest advantage of *in silico* technologies. As an alternative to generic parallelization on commodity clusters, it has recently been shown that specialized, purposed built computer architecture can accelerate MD calculations by two orders of magnitude [33].

Conclusion and future works

An integrated SSBDD platform has been built to create active novel chemical structures for chemical synthesis. A novel feature of the platform is its ability to calculate ligand affinity constants with great accuracy. Despite the substantial computational resources required to achieve this accuracy, the approach is cost effective because it dramatically reduces the time and expense associated with unnecessary experimental synthesis and assays, enabling chemists to focus their synthetic efforts chemical structures that are most likely to lead to therapeutic success. Currently, binding affinity prediction with MAPLE CAFEE tends to be unstable when the actual affinity is very weak or ligands are protonated. We are now improving the sampling algorithms for weak binding state, and charge correction methods for protonation. We are also reducing the requisite CPU resources as much as possible without sacrificing accuracy.

The availability of next generation super-computers puts a practical thermodynamic guided drug-design platform,

(Figure 3 Legend Continued) abstract fragments following drug-design strategies. (c) Exhaustively connect the selected fragments with constraints to generate as many abstract molecular skeletons as possible, then stabilize them with MM. (d) Assign possible real-atoms to the abstract skeletons considering biophysical constraints and synthetic feasibility. (e) Screen out unfavorable structures using heuristics such as the occurrence of pharmacophores that raise toxicological 'red flags'. The standard MAPLE CAFEE workflow comprises five steps (f)–(j). (f) Setup molecular models for protein, ligand, and water and assign the force field parameters with Force Field Formulator for Organic Molecules (FF-FOM) [28]. When no co-crystal is available, reasonable initial binding poses should be obtained with docking or other methods. It is notable that having realistic molecular models with refined force field parameters is essential as well as efficient algorithms with reliable computational method for estimation. (g) Equilibrate the system in the initial pose with a MD job. (h) Compute the nonequilibrium work term between bound and unbound equilibrium states. Actually, the two end states are divided into microstates with decoupling parameters, and the trajectories for each microstate are computed with MD jobs. (i) Estimate the most likely representative binding free energies between microstates with the BAR (Bennett Acceptance Ratio) statistical method [34]. (j). Finally, sum up all the micro binding free energies to get the total binding free energy. Molecules can be submitted to MAPLE CAFEE for free energy calculation from any source. For example, they might originate as proposed molecules in a medicinal chemistry series (k), or from a DOCK type virtual screen (l).

both accurate and swift, within reach. Moving forward, accelerated computation, and the virtual drug discovery platforms it will support, will bolster successful therapeutic R&D outcomes and contribute to better human health.

Conflict of interest statement

SM is General Manager of the *in silico* Drug Discovery Research Division, Bio-IT Business Development Unit, Fujitsu Limited, which has a commercial interest in MAPLE CAFEE, OPMF, and the SSBDD PLATFORM. WM collaborates with the Fujitsu group but has no financial interest. WM is supported by a block grant to the Experimental Therapeutics Centre from the Biomedical Research Council of the Agency for Science and Technology Research (A*STAR), Ministry of Trade and Industry, Singapore.

Acknowledgements

SM would like to express his gratitude to all the researchers for their theoretical and practical efforts on free energy computation and also engineers concerned with super-computer. Dr. Hirofumi Watanabe in Kobe University (currently Riken) organized a project for Case-1 and Dr. Makoto Taiji and his group in Riken willingly participated in it. Dr. Yasuhiko Shiratori and his group in Chugai Pharmaceutical Co. Ltd. configured the blind-test for Case-2 and re-measured the binding affinity for validation. WM thanks the Institute of High Performance Computing and the A*STAR Computational Resource Centre for their support in advancing computational molecular design in Singapore.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- of special interest
 - of outstanding interest
1. Kaitin KI, DiMasi JA: **Pharmaceutical Innovation in the 21st century new drug approvals in the first decade, 2000–2009.** *Clin Pharmacol Ther* 2011, **89**:183-187.
 2. Adams CP, Van Vu Brantner VV: **Spending on new drug development.** *Health Econ* 2010, **19**:130-141.
 3. David E, Tramontin T, Zimmel R: **Pharmaceutical R&D: the road to positive returns.** *Nat Rev Drug Discov* 2009, **8**:609-610.
 4. Boucher HW, Talbot GH, Bradley JS, Edwards JE, Gilbert D, Rice LB, Scheld M, Spellberg B, Bartlett J: **Bad bugs, no drugs: no ESCAPE! An update from the Infectious Diseases Society of America.** *Clin Infect Dis* 2009, **48**:1-12.
 5. Waterer G: **Controlling epidemic viral infection.** *Curr Opin Infect Dis* 2011, **24**:130-136.
 6. DeVol R, Bedrouslan A, Charowron A, Chatterjee A, Kim LK, Klowden K: *An Unhealthy America: Charting a New Course to Save Lives and Increase Productivity and Economic Growth.* Miliken Institute; 2007.
 7. David E, Tramontin T, Zimmel R: **The road to positive returns.** *McKinsey Q* 2010, **February**:1-5.
 8. Pauling L, Delbruck M: **The nature of the intermolecular forces • operative in biological processes.** *Science* 1940, **92**:77-79. The fundamental insight is '...regarding complementarity... Attractive forces between molecules vary inversely with the power of the distance, and maximum stability of a complex is achieved by bringing the molecules as close together as possible in such a way that positively charged groups are brought near negatively charged groups... minimum distances of approach of atoms are determined by the their repulsive potentials, which may be expressed in terms of van der Waals radii; in order to achieve maximum stability, the two molecules must have complimentary surfaces, like die and coin and a complimentary distribution of active groups'.
 9. Kuntz ID, Blaney JM, Oatley SJ, Langridge R, Ferrin TE: **A geometric approach to macromolecule–ligand interactions.** *J Mol Biol* 1982, **161**:269-1288.
 10. Gund P, Andose JD, Rhodes JB, Smith GM: **Three-dimensional molecular modeling and drug design.** *Science* 1980, **208**:1425-1431.
 11. Ripphausen P, Nisius B, Peltason L, Bajorath J: **Quo Vadis, virtual •• screening? A comprehensive survey of prospective applications.** *J Med Chem* 2010, **53**:8461-8467. A meta-analysis of 2995 virtual screening studies, examining the types of methods employed, their rates of success, the preferred targets, and the potency of compounds identified. Only papers with accompanying experimental evaluation of the compounds were considered. The data set was drawn from Journal of Medicinal Chemistry, Bioorganic Medicinal Chemistry Letters, Bioorganic Medicinal Chemistry, Journal of Chemical Information and Modeling, ChemMedChem, European Journal of Medicinal Chemistry, Chemical Biology and Drug Design, Journal of Computer-Aided Molecular Design, ACS Chemical Biology, ChemBioChem, Nature Chemical Biology and Angewandte Chemie (International Edition). Greater than 99% of the papers appeared after 2000, and greater than 80% after 2005. Enzymes represented the largest target class (with kinases, proteases and phosphatases being the best represented, in that order). Other popular targets were membrane receptors, and transcription factors. Structure based methods were preferred to ligand based methods by a ratio of 3:1, with hit compounds produced by each method in about the same ratio. However, the structure based approaches found hits across a broader distribution of potencies, while hits from the ligand based methods were concentrated in most potent category (<1 mM). Interestingly, measured by the number of sub mM hits discovered, docking into homology modeled structures slightly outperformed docking into physically determined structures (X-ray or NMR).
 12. Joachimiak A: **High-throughput crystallography for structural genomics.** *Curr Opin Struct Biol* 2009, **19**:573-584.
 13. Irwin JJ, Shoichet BK: **ZINC – a free database of commercially available compounds for virtual screening.** *J Chem Inf Model* 2005, **45**:177-182.
 14. Taylor P, Blackburn E, Sheng YG, Harding S, Hsin K-Y, Kan D, Shave S, Walkinshaw MD: **Ligand discovery and virtual screening using the program LIDAEUS.** *Br J Pharmacol* 2008, **153**:555-567.
 15. Blum LC, Reymond J-L: **970 Million drug like small molecules for virtual screening in the chemical universe database GDB-13.** *J Am Chem Soc* 2009, **131**:8732-8733.
 16. Perola E, Walters W, Charidson PS: **A detailed comparison of current docking and scoring methods on systems of pharmaceutical relevance.** *Proteins: Struct Funct Bioinform* 2004, **56**:235-249.
 17. Cheng T, Li X, Li Y, Liu Z, Wang R: **Comparative assessment of •• scoring functions on a diverse test set.** *J Chem Inf Model* 2009, **49**:1079-1093. 16 Scoring functions assessed on 195 protein–ligand complexes rationally selected from PDB bind with high resolution crystal structures and experimentally determined binding constants. This reduced set was chosen from representative members of clusters from a larger 1300 protein–ligand ensemble from PDB bind. The reduced set included 65 different protein families, with a high affinity, a medium affinity and a low affinity pair selected to represent each cluster. In addition the four most frequently represented families were retained in mass to assess scoring performance with related clades. The scoring functions were evaluated for: first, docking power, 'the ability to identify the true ligand binding pose' (metric: root-mean-square deviation from crystal structure <2 Å), second, ranking power, 'the ability to correctly rank different ligands bound to the same receptor when the correct binding poses ... are known' (metric: concordance of computational scores and experimental binding affinities within the 65 proteins families in the test set, three protein–ligand complexes per family, having high, medium and low affinity, i.e. if the computational prediction ranked them in correct order then it was scored as a success, if any other of the 5/6 possible orders was predicted, then it was scored as a failure; also Spearman and Pearson statistics on concordance within the other four highly represented families), and third, *scoring power*, 'the ability of calculating binding scores that correlate with experiment' across a biophysically diverse spectrum of protein–ligand pairs (metric: Spearman and Pearson

correlations on the base set). Authors conclude that 'no single scoring function consistently outperforms the others in all aspects' and recommend that dockers 'choose the appropriate scoring functions for different purposes'. Scoring functions evaluated are: X-score (version 1.2), DrugScore version 1.2 and online version 0.9, GlideScore version 4.5, three scoring functions from GOLD version 3.2 (GOLD::GoldScore, GOLD::ChemScore, and GOLD::ASP), five from Discovery Studio version 2.0 (DS::LigScore, DS::PLP, DS::Jain, DS::PMF, and DS::Ludi), and five from SYBYL version 7.2. Output from the various platforms was adjusted so as to make comparison of the results meaningful. In addition to the author's own analysis, the supplementary material offers a detailed summary of DOCK evaluations, published since 2000.

18. Plewczynski D, Lazniewski M, Augustyniak R, Ginalski K: **Can we trust docking results? Evaluation of seven commonly used programs on PDBind database.** *J Comput Chem* 2011, **32**:742-755.
Three categories of DOCKing software were evaluated: genetic algorithmic (GOLD and AutoDock), fragment based (FlexX and Surflex) and geometrical, that is, shape-complementarity based (LigandFit and Glide). The test set was 1300 ligand-protein pair structures from PDBbind that have accompanying, experimentally determined binding data. The metric used was the Spearman's rank correlation of experimentally measured binding and computational binding score.
19. Wang R, Lu Y, Wang S: **Comparative evaluation of 11 scoring functions for molecular docking.** *J Med Chem* 2003, **46**:2287-2303.
20. Kontoyianni M, McClellan LM, Sokol G: **Evaluation of docking performance: comparative data on docking algorithms.** *J Med Chem* 2004, **47**:558-565.
21. Warren GL, Andrews CW, Capelli A-M, Clarke B, LaLonde J, Lambert MH, Lindvall M, Nevins N, Semus SF, Senger S *et al.*: **A critical assessment of docking programs and scoring functions.** *J Med Chem* 2006, **49**:5912-5931.
22. Cummings MD, DesJarlais RL, Gibbs AC, Mohan V, Jaeger EP: **Comparison of automated docking programs as virtual screening tools.** *J Med Chem* 2005, **48**:962-976.
23. Jarzynski C: **Nonequilibrium equality for free energy differences.** *Phys Rev Lett* 1997, **78**:2690.
24. Jarzynski C: **Equilibrium free-energy differences from non equilibrium measurements: a master-equation approach.** *Phys Rev E* 1997, **56**:5018.
25. Jorgensen WL, Buckner JK, Boudon S, Tirado-Rives J: **Efficient computation of absolute free energies of binding by computer simulation. Application to the methane dimer in water.** *J Chem Phys* 1988, **89**:3742.
26. Shirts MR, Bair E, Hooker G, Pande VS: **Equilibrium free energies from nonequilibrium measurements using maximum-likelihood methods.** *Phys Rev Lett* 2003, **91**:140601.
27. Fujitani H, Tanida Y, Ito M, Jayachandran G, Snow CD, Shirts MR, Sorin EJ, Pande VS: **Direct calculation of the binding free energies of FKBP ligands.** *J Chem Phys* 2005, **123**:084108.
28. Fujitani H, Matsuura A, Sakai S, Sato H, Tanida Y: **High-level ab initio calculations to improve protein backbone dihedral parameters.** *J Chem Theory Comput* 2009, **5**:1155-1165.
29. Tanida Y, Ito M, Fujitani H: **Calculation of absolute free energy of binding for theophylline and its analogs to RNA aptamer using nonequilibrium work values.** *Chem Phys* 2007, **337**:135.
30. Fujitani H, Tanida Y, Matsuura A: **A massively parallel computation of absolute binding energy with well-equilibrated states.** *Phys Rev E* 2009, **79**:021914.
31. Shaw DE, Deneroff MM, Dror RO, Kuskin JS, Larson RH, Salmon JK, Young C, Batson B, Bowers KJ, Chao JC *et al.*: **Anton, A special-purpose machine for molecular dynamics simulation.** *Commun ACM* 2008, **51**:91-97.
32. Watanabe F, Tanaka S, Okimoto N, Hasegawa A, Tajji M, Tanida Y, Mitsui T, Katsuyama M, Hideaki F: **Comparison of binding affinity evaluations for FKBP ligands with state-of-the-art computational methods: FMO, QM/MM, MM-PB/SA and MP-CAFFE approaches.** *Chem-Bio Informatics J* 2010, **10**:32-45.
33. Klepeis JL, Lindorff-Larsen K, Dror RO, Shaw DE: **Long-timescale molecular dynamics simulations of protein structure and function.** *Curr Opin Struct Biol* 2009, **19**:120-127.
34. Bennett CH: **Efficient estimation of free energy differences from Monte Carlo data.** *J Comput Phys* 1976, **22**:245.