Published in final edited form as:
Curr Top Med Chem. 2010 ; 10(1): 3–13.

Emerging Methods for Ensemble-Based Virtual Screening

Rommie E. Amaro[1] and **Wilfred W. Li**[2]

[1] Department of Pharmaceutical Sciences and Department of Information and Computer Science, University of California, Irvine; Irvine, CA 92697

[2] National Biomedical Computation Resource, University of California, San Diego, La Jolla, CA 92093

Abstract

Ensemble based virtual screening refers to the use of conformational ensembles from crystal structures, NMR studies or molecular dynamics simulations. It has gained greater acceptance as advances in the theoretical framework, computational algorithms, and software packages enable simulations at longer time scales. Here we focus on the use of computationally generated conformational ensembles and emerging methods that use these ensembles for discovery, such as the Relaxed Complex Scheme or Dynamic Pharmacophore Model. We also discuss the more rigorous physics-based computational techniques such as accelerated molecular dynamics and thermodynamic integration and their applications in improving conformational sampling or the ranking of virtual screening hits. Finally, technological advances that will help make virtual screening tools more accessible to a wider audience in computer aided drug design are discussed.

INTRODUCTION

The computational identification of drug leads out of large compound libraries through receptor-based virtual screening (VS) is a well-established method to predict putative inhibitors for target receptors (reviewed in [1–9]). Despite advances in the underlying algorithms of virtual screening experiments have incrementally improved our ability to discriminate binders from non-binders, a number of obstacles remain. A major outstanding challenge in the practice of virtual screening is the treatment of receptor flexibility, which is especially difficult owing to the many degrees of conformational freedom in target receptors. Steady increases in computational power, coupled with improvements in the underlying algorithms and available structural experimental data, are enabling a new paradigm for virtual screening, wherein computationally predicted ensembles from first-principle simulations are being used in rational drug design efforts. The integration of these more rigorous physics-based methods will have far reaching impact on translational medicine, including the ability to: (1) better understand the structural dynamics of disease-related target receptors, (2) improve our quantitative assessments of ligand-receptor interactions, (3) discover novel modes of ligand binding and inhibition, and (4) develop new therapeutics that are patient-specific and less prone to drug resistance. This review will attempt to summarize the recent computational and theoretical advances that have enabled the development of new methods to integrate larger-scale sampling of receptor space and identify the most biologically relevant structures for drug design. We will attempt to discuss the advantages and drawbacks of current methodologies, but focus on emerging methods

*Address correspondence to this author at the Department of Pharmaceutical Sciences and Department of Information and Computer Science, University of California, Irvine; Irvine, CA 92697.

that we believe will become more frequently utilized, especially for receptors that exhibit a high degree of flexibility.

MOVING BEYOND "LOCK-AND-KEY" OR SIMPLE "INDUCED FIT"

A full understanding of molecular recognition presents a problem of intense interest to molecular sciences and the field of computer-aided drug design (CADD) [10–13]. The interactions between ligand molecules and their corresponding receptors are dynamic and complex. Over time, the field has moved away from an original understanding of molecular recognition as a simple, rigid "lock-and-key" mechanism [14], towards an "induced fit" [15]. In the traditional induced-fit theory, the intrinsic plasticity of the receptor is utilized when the protein's structure and function responds to the ligand-binding event through inducible structural motion, and the corresponding rearrangements of the receptor active site upon binding are only capable in the presence of the bound ligand. Many studies describing the induced fit effects in various protein-ligand systems have been presented (for examples, see [16–21]). It is widely accepted that ligands may bind to receptor conformations that occur infrequently in the receptor's dynamics, and that the local motions of active site residues can drastically alter the binding affinity and specificity of ligands to their target. The ability to efficiently sample these rare dynamics and furthermore, to incorporate the resulting conformations into the drug discovery and design protocol is still an active area of investigation. Instead of using static crystal structures in virtual screening, in the traditional lock and key sense, methods that explore large domain motions and significant active site rearrangements in a computationally efficient manner will be required.

CONFORMATIONAL ENSEMBLES

An important theoretical notion that presented a modification to traditional induced fit theory is that of ligands binding to pre-existing receptor populations, or conformational ensembles [22, 23]. Borrowed in large part from the more theoretically-leaning field of protein folding, the general idea is that the receptor does not exist in a single native conformation, but more realistically exists in an ensemble of metastates, or energetically low-lying substates along the rugged receptor energy landscape [24, 25]. The ligand is exposed to this ensemble of receptor conformational states and may bind preferentially to one, shifting the equilibrium population towards the favorable bound conformation [26]. These theories are deeply rooted in statistical mechanics and state that the ensemble comprises a statistical distribution of native conformations that essentially move, subject to mechanical forces. Although a review of the elegant protein folding theories is outside the scope of this review (for comprehensive reviews, see [24, 27, 28]), it is important to recognize the influence of these theories on understanding the fundamental nature of protein structure, function and ligand binding [29]. An essential conceptual point that is highly relevant to this review is the recognition that receptors are able to sample productive bound-ligand conformations even while in the apo state. To some extent, this suggests that the "lock and key" is but one of the rare conformations within this unifying scheme, and that conformational selection is an important driving force in ligand binding and recognition.

Using multiple receptor conformations in CADD and virtual screening ensued not long after the conception of conformational ensembles. These receptor conformations may be obtained from x-ray crystallography, or nuclear magnetic resonance (NMR) experiments, or computationally, from molecular dynamics (MD) or Monte Carlo simulations. In a critical advance, Knegtel *et al*. [30] presented the first study employing both crystallographically and NMR derived ensembles in an averaged grid-based docking study. Other early and important examples of utilizing multiple experimental structures include FlexE [31], which systematically combined features from multiple crystal structures, and work presented by

Osterberg *et al.*, which carefully examined different methods to combine receptor grids stemming from the inclusion of almost two dozen receptor structures of HIV-1 protease [32]. Numerous additional studies continue to enhance the notion that using multiple receptor conformations improves enrichment factors and the ability to predict viable binding modes [22, 33–38].

Although experimentally derived structures have typically been considered the gold standard for docking and drug design, in many cases, a high correspondence between computationally derived ensembles and experimental structures has been established, though initially limited by insufficient sampling [39]. The progress to date suggests that MD-generated ensembles may closely replicate the structural dynamics of proteins in solution, as shown by NMR experiments [40, 41]. With this in mind, we focus on emerging virtual screening methods that utilize direct sampling of receptor conformational space through MD-based approaches, and note that comprehensive reviews of other methods to include receptor flexibility, including soft docking [42], sidechain rotamer libraries, docking to relevant normal modes [43], induced fit docking [44], and more are presented elsewhere [45–47].

COMPUTATIONALLY GENERATED ENSEMBLES FOR DRUG DISCOVERY

Dynamic Pharmacophore Model

Although the first MD simulation of a protein was performed over 30 years ago [48], the more systematic use of the resulting conformational ensembles in a predictive fashion has only recently been successfully demonstrated. The first experimentally verified study to use multiple computer-generated structures in a systematic ensemble-based approach to discovery was the multiple protein structures based dynamic pharmacophore (DPM) model approach presented by Carlson *et al*. [49]. In this work, a 500 ps explicitly solvated MD simulation of the apo HIV-1 integrase catalytic domain was performed [50] using the AMBER95 force field [51] and the NWChem v3.2 simulation program [52]. The flexible nature of the integrase system in the active site region precluded resolving the structure completely crystallographically; instead, the authors used a homology model (modeled after the complete structure of Avian Sarcoma Virus integrase) of the flexible region near the active site. Their predicted structure (and the corresponding MD trajectory) was later validated when two additional crystal structures of the full integrase catalytic domain were published, showing a high degree of similarity between the predicted regions and the new crystal information [49]. Eleven conformations from the MD trajectory were used as the representative ensemble to create a dynamic receptor-based pharmacophore model for the integrase system. The bare active sites were flooded with small molecule probes representing different chemical functional groups, which essentially map the most favorable areas in the active site, and are clustered across the ensemble of structures. The resulting DPM was used to search the Available Chemical Database for potential inhibitors, and rerank the known set of active compounds for the integrase. Experimental verification of the predicted set indicated that about one third of the compounds were inhibitory and the reranking with the DPM outperformed any single static pharmacophore.

Refinements to the DPM method were presented in an application focusing on the flexible HIV-1 protease as a test case [53]. In this work, Meagher and Carlson showed improved performance of the DPM to discriminate binders from non-binders with increasing lengths of MD simulations. This result reinforced not only the importance of including receptor flexibility in CADD efforts, but also the consideration that longer simulation lengths yield improved results for discovery. Presumably, the longer simulation times enhanced the conformational sampling of receptor energy landscape.

More recently, the DPM was used to discover novel inhibitors of the MDM2-p53 protein [54]. Snapshots extracted from a 2 ns MD simulation of the human MDM3 bound to p53 were extracted every 100 ps. These 21 structures were used to create a 6-site pharmacophore model of the active site. A virtual screening of 35,000 in-house compounds identified 27 compounds, of which 23 were experimentally tested. Four active compounds with unique chemical scaffolds were discovered to inhibit at 50 μM concentration or better. Interestingly, through a close analysis of the MD simulations and crystal structures, the authors had discovered an additional hydrophobic binding area near the known binding cleft [55]. Creating an additional pharmacophore model that utilized this site resulted in the discovery of a fifth compound that inhibited at ~ 20 μM. The discovery of these new compounds is an important step forward in the use of MD-generated structures in virtual screening, and validation of the dynamic pharmacophore method.

The Relaxed Complex Scheme

Another approach of using MD-generated structural ensembles for drug discovery utilizes a strategy referred to as the *relaxed complex scheme* (RCS) (Fig. 1). The RCS uses receptor snapshots extracted from MD simulations to search ligand libaries via small molecule docking [56, 57]. It combines the advantages of docking algorithms with dynamic structural information provided by MD simulations, explicitly accounting for the flexibility of both the receptor and docked ligands. This procedure directly applies the concept of conformational ensembles in molecular recognition through docking. Increasingly longer time scale MD simulations enhances our ability to effectively sample the receptor conformational space prior to docking. This scheme has been developed in combination with various MD software packages and AutoDock for the ligand docking [58], although other docking programs can be substituted. The RCS was first applied to the FKBP binding protein [59] and tested using improved re-scoring functions based on MM-PBSA models [56].

An important example of the method was presented in Schames *et al*., who performed a 2 ns explicitly solvated MD simulation of HIV-1 integrase in complex with the 5CITEP inhibitor [60]. The simulations revealed an additional cavity adjacent to the integrase active site (i.e. a “trench”), and docking of 5CITEP into this new area proved to be even more favorable than the actual active site based on the AutoDock 3.0 scoring function. Shortly thereafter, Hazuda *et al*. acknowledged that this new structural understanding, in which essential interactions occur in varying areas of the active site, was invaluable in creating new integrase inhibitors with unique resistance profiles [61]. Just a few years (and rounds of optimization) later, raltegravir was approved for use in humans [62]. This success story provides renewed motivation for the development of computational methodologies that incorporate predictive structural information for flexible receptors systematically.

ENSEMBLE SELECTION

In order to make these more rigorous yet time-consuming computational methods accessible to a wider audience, various ways to distill the structural ensemble, and thus reduce the requisite computational power, need to be explored. With constant improvements in computing power and the underlying simulation algorithms, we will very soon routinely be sampling into the microsecond timescale [63]. Performing virtual screening experiments on the full set of resulting structures is computationally intractable and likely unnecessary. New metrics to reduce the ensemble systematically – without losing critical structural information – will be important. Several strategies have been developed that select structural information from the resulting ensemble, for both the DPM and the RCS. These approaches, which range from manual selection to more systematic mathematical techniques, both reduce the ensemble to a manageable size and allow the extraction of the most meaningful (i.e. biologically relevant) information.

RMSD-BASED CLUSTERING

RMSD-based clustering has been established as an alternate metric that can be used in order to reduce an ensemble to a meaningful set before virtual screening experiments. RMSD-based clustering can be especially useful because it provides information about the most dominant configurations sampled during an MD trajectory. One can argue based on statistics that the most populated clusters of receptor structures would be more frequently encountered by the ligand, energetically stable, and statistically meaningful. Methods remain to be developed that could incorporate both population and energy states, when seeking to create more minimal representative ensembles from MD trajectories.

Deng *et al*. presented a refinement of the DPM that combined RMSD-based clustering in a hierarchical fashion with energetic probabilities in order to intelligently select which structures to include in the pharmacophore model [64]. In this study, the authors performed a 1 ns simulation of the HIV integrase in complex with the 5CITEP inhibitor, again using homology modeling to complete part of the active site structure [65]. 1000 snapshots were extracted from the MD trajectory at 1 ps intervals. RMSD-based clustering based on five key active site residues was performed, followed by a relative estimation of the probability of each MD snapshot based on a Boltzmann factor ($e^{-\Delta U/RT}$), wherein the energy difference was based on the difference from the minimum energy structure [66]. The cluster with the largest total sum of the Boltzmann factors was selected as the first cluster. The snapshots from this first cluster were removed from the ensemble and the process repeated to determine the second cluster, and so forth. The final computational ensemble comprised 10 receptor conformations, and collectively represented approximately 50% of the trajectory. LigBuilder [67] was then used to create a DPM based on these 10 representative cluster structures and Catalyst 4.6 [68] was used to search an in-house database of 400 available chemical compounds. A final set of 23 compounds was selected for experimental assays and 9 showed inhibitory capacity at 100 μM or less. Several compounds so identified were missed if only a static pharmacophore model was used.

Cheng *et al*. recently extended the RCS for virtual screening by the efficient use of RMSD-based clustering information as well [69] In order to distill the most dominant configurations from the MD simulations, Cheng *et al*. performed RMSD-based clustering on snapshots extracted every 10 ps from two 40 ns trajectories of avian influenza N1 neuraminidase in the apo form and in complex with the inhibitor oseltamivir [70]. Although the tetramer N1 was used in the simulations, the clustering analyses were carried out on individual monomer protein chains [71]. Therefore, each 40 ns tetramer simulation yielded the equivalent of four times the monomer sampling (160 ns), yielding 16,000 structures for the analyses. To date, this study was based on considerably longer MD trajectories than any other published examples used in CADD (the equivalent of 160 ns for the monomer for each of the apo and holo systems). Therefore an extensive sampling of the receptor configurational space was achieved; in addition, the four monomer copies in the tetramer allows for a multi-copy MD approach, which has been shown to enhance sampling as opposed to a single long trajectory [72]. The RMSD-clustering was performed on a subset of binding-site residues and yielded 10 structures representing the apo ensemble and 5 structures representing the holo ensemble. The central member structure from the three most dominant clusters in the apo and holo sets were screened against the National Cancer Institute Diversity Set 1 (NCIDS1) using AutoDock 4.0 [73]. The final ranking of all ligands from 8 primary screens was determined by taking the weighted average of the docking scores into the full representative ensemble of the holo MD trajectory (Fig. 2). Of the experimentally tested set of 25 compounds, 10 exhibited K_i's of under 500 μM. Of these, 7 compounds were only selected through the screening of the MD-generated structures, which exhibited large structural rearrangements during simulation (Fig. 2) [74]. Eight of the compounds are predicted to at least partially

bind to novel binding sites revealed in the simulations. This significant reordering enabled the identification of hits that otherwise would have been missed.

In a separate system published at almost the same time, Zhong *et al*. performed an explicitly solvated 5 ns MD simulation of human DNA ligase 1 using the CHARMM27 force field [75] with CMAP corrections [76] and the CHARMM MD engine [77]. Receptor coordinates were extracted every 5 ps (i.e. 1000 conformations total) [78]. They then performed RMSD-based clustering on binding site residues and used representative structures from the four largest clusters in a second round of docking. The top 50,000 compounds from the first phase were docked using DOCK4.0 [79] into the crystal structure as well as representative structures from the five most populated clusters. Compounds were ranked according to the most favorable individual score from the set and 192 compounds tested experimentally. Of the 10 active compounds discovered, nine were chosen based on MD-generated representative clusters.

QR-FACTORIZATION

An alternate technique that has been applied to virtual screening experiments using the RCS is QR-factorization. The technique was originally designed to remove inherent bias in structure databases and distill, from a vast quantity of redundant information, a minimal basis set of protein structures accurately spanning the evolutionary conformation space of a particular protein [80]. In 2008, Amaro *et al*. incorporated this new method to create a minimal basis set for the configurational space sampled in the MD simulations and used the resulting ensemble in a virtual screening with the RNA editing ligase 1 enzyme in *Trypanosoma brucei* [81]. In this work, a 20 ns explicitly solvated MD simulation employing the Charmm27 force field [75] and the NAMD2.6 MD program [82] were performed of the ATP-bound enzyme [83]. Snapshots of the receptor only (ATP and all water molecules removed) were extracted every 50 ps, generating a total of 400 receptor conformations. The QR-factorization algorithm reduced the initial set of 400 MD structures to 33. In the first round, virtual screen of the NCIDS1 was performed using the static crystal structure and AutoDock 4.0 [73] to dock and rank approximately 1800 compounds. In order to take receptor flexibility into account and to validate and refine the top hits, the top 2% of the screening hits (corresponding to the top thirty compounds), were redocked into the full and QR-reduced holo MD-ensembles. A comparison of the mean RC energy (i.e. mean energy from docking into all 400 snapshots) and the mean QR-RC energy (i.e. mean energy from docking into the representative 33) for the top thirty compounds indicated that redocking into the QR-reduced set was a more efficient way to capture the effects of receptor flexibility without loss of energetic information. In total, 10 compounds were experimentally tested and 5 were found to inhibit at 10 μM or better. Importantly, two of the best inhibiting compounds would not have been selected for testing based on the crystal structure score alone, therefore, further refinement of the predicted compounds based on the RCS reranking provided an important enrichment of the final ranked set.

These studies clearly highlight the significance of including MD-generated conformations in the virtual screening protocol for discovering new actives. The identification of entirely novel modes of binding in the hit identification stage provides a key advantage over other methods. This is especially true for very flexible receptors, in which entirely new pockets may be revealed for potential ligand binding. The success of these predictions underscores the importance of incorporating receptor flexibility in docking and scoring methods used in conjunction with virtual screening.

CURRENT CHALLENGES AND OPPORTUNITIES

Knowledge Gaps

One issue facing many of the more intensive physics-based methods is the knowledge gap between industry and academia, as well as computational and experimental academic groups. Despite an increasing number of academic groups working in CADD, there is still a knowledge gap between industrial and academic groups working in the field. Similarly, computational academic groups do not always publish with experimental validation (in this review, we attempted to only cover computational approaches for which experimental verification was available). The interdisciplinary nature of drug discovery begs for a close interaction between all of these groups. It will be important to leverage knowledge from both camps going forward if we are to maximize progress in improving the "state-of-the-art." In this regard, building bridges that allow the field to utilize data and insight from industry in conjunction with new computational and theoretical methodologies being developed in academia will be especially useful. Additionally, computational and theoretical academicians who are pursing these more rigorous and innovative drug discovery approaches must work more closely with experimentalists to validate their methods. More extensive academic and industrial collaborations in the area of drug discovery for neglected infectious diseases will also be progressively more important given the present climate changes and an increasingly mobile world population.

Predicting Explicit Water Interactions

Another current methodological challenge in virtual screening is the inclusion of explicit ordered water molecules. Water is well-known to play an important role in ligand binding, either through the formation of key lubrication contacts between the receptor and ligand, or by being displaced upon ligand binding [84]. A high percentage of high-resolution crystal structures contain resolved water molecules in or near the active site [85], but the decision as to whether or not to retain them in virtual screening experiments is unclear [86]. A recent study sampling the effects of including different water molecules in docking calculations for 24 receptor targets indicated improved enrichment for approximately half of the receptors studied [87]. A complicating consideration is the potential introduction of hydration artifacts due to the cryogenic temperatures used to trap many receptor structures [88]. MD simulations that include experimentally resolved water molecules can provide easily accessible additional information about these water molecules, such as residence times and dipole order parameters; this information could be used to determine whether to include specific water molecules in virtual screening experiments for a given receptor. More accurate energetic approaches, such as calculating the free energy of binding for individual water molecules can also be considered [89]. Future MD-based ensemble approaches will hopefully make use of the structural water information to find novel candidate compounds in virtual screening strategies.

Promising New Ways to Enhance Sampling of the Receptor Ensemble

The ability to sample longer timescales presents yet another challenge for traditional all-atom MD simulations, despite continuous increases in computer power. Sampling routinely into the microsecond timescale and beyond will require the use of new and improved techniques. New and promising emerging approaches that address the general problem of more efficiently sampling receptor conformational space are being pursued; two approaches in particular that we would like to highlight that have benefitted recently from a number of applications on more realistic protein systems are accelerated molecular dynamics (AMD) and generalized Born MD (GB MD). Through the use of a robust potential function, AMD increases the transition of high-energy barriers without any advance knowledge of the underlying energy landscape [90]. This promising method has been shown to sample slow

diffusive conformational transitions [91] as well as speed rates of convergence for entropic calculations [92]. Importantly, experimental validation of accurate, enhanced conformational sampling through AMD has been carried out by NMR studies [93]. We anticipate that the structures sampled with this far-reaching technique will facilitate the discovery of new structural understandings that will lead to the discovery of novel therapeutics, and also potentially provide more accurate estimates of the thermodynamics of binding. A second technique to pave the way into the microsecond regime that is ready to be applied to larger biomolecular systems, is generalized Born molecular dynamics (GB MD). GB MD, which employs a continuum representation of the solvent water molecules and salt ions, offers computational efficiency over the more rigorous Poisson-Boltzmann solvers or explicit solvent simulations [94]. The implicit solvent approach essentially reduces solvent friction and has been shown to enhance conformational sampling for a wide variety of peptide, protein, and nucleic acid systems [95–97], and more recently, larger protein systems [98, 99]. Correspondence between GB models and experimentally derived structures has also been established [100, 101], which, similar to AMD, makes it a good choice for extensions of the computational receptor ensemble.

How Long is Long Enough?

A key question regarding the MD-based approaches for conformational ensemble generation is determining for how long one should simulate. Unfortunately there does not yet appear to be a clear-cut answer to this question. Current successful published studies range from 500 ps to 320 ns in length, which is quite a broad range of timescales. Each of the studies presented some measure of success, mainly the identification of compounds that otherwise would have been missed in static receptor models. One could determine enrichment factors as a function of simulation length for a particular receptor, although it would be difficult to generalize the findings to all possible receptors since different proteins exhibit varied structural dynamics. Moreover, rare events that occur on longer-time scales will only be sampled if the simulations are extended beyond the requisite time point. Including conformational ensembles generated with accelerated sampling approaches, such as AMD or GB MD, may shed light on this aspect. For now, the only clear answer seems to be that including receptor flexibility to some degree, even if only explored for 1 ns of dynamics, is better than a purely rigid model.

Application of more Rigorous Approaches for Predicting Binding Affinities

Although in virtual screening experiments success is more loosely defined as being able to separate the binders from non-binders, a challenge intimately connected to virtual screening is the development of methods to predict the binding affinities for a series of compounds to a target receptor (reviewed in, among others: [102–107]). Most current methods apply the indirect approach first suggested by Tempe and McCammon, which utilizes a simple thermodynamic cycle for the bound and unbound receptor-ligand complex [108]. The most rigorous treatment of receptor-ligand binding that presently exist are the so-called “computational alchemy” techniques achieved through MD simulations, in which free energy changes are estimated based on coupling-parameter approaches, such as free energy perturbation or thermodynamic integration [109]. These methods describe a higher physical complexity of the binding process and include an extensive sampling of receptor, ligand, and solvent phase spaces 1[07, 110–119]. Importantly, these methods provide more accurate estimates of binding free energies as well as reliable measures of their accuracy. Although they suffer from extraordinary computational costs that essentially prohibit their mainstream application, this too, is changing. More frequently these are being applied to larger ligand sets, and although computational power has not advanced to the point of being able to apply these methodologies to entire ligand libraries, they are being used to handle several dozens of candidate compounds [120]. Additionally, these algorithms now exist as functional

modules in the massively parallelizable MD programs Desmond (TI) [121] and NAMD (FEP) [114], allowing the determination of accurate free energies of binding in equivalent wall-clock time of hours-to-days, whereas serial implementations of these algorithms traditionally require(d) weeks-to-months of computing time for a single compound. The utilization of petascale computing infrastructures that will soon become publicly available will continue to drive down the cost of such calculations. Faster and more approximate free-energy based methods, such as molecular mechanics Poisson Boltzmann (generalized Born) surface area (MM-PB(GB)SA) [122, 123], linear interaction energy [124, 125], single-step perturbation [126–129], are less computationally intensive and thus already more easily extendible to larger compound libraries.

CLOUD COMPUTING, GPU AND HARDWARE ACCELERATION

The increasing availability of computing power on the order of peta- to exa-FLOPS (floating point operations per second) will make possible longer time course simulations (Fig. 3). Currently, a 100 ns explicit solvent simulation of a 75 kDa protein can be completed in one week on a supercomputer or a local cluster with low latency network. More rigorous physics-based approaches are increasingly possible. The arrival of supercomputers such as the NCSA/IBM Blue Waters with many-core architecture will enable simulations on the order of 1~100 μs on a routine basis. The advance in general purpose graphics processing unit (GPGPU) based processing power, where each GPU contains as many as 512 cores, means that the age of personal supercomputer has dawned upon us. While GPGPU is still challenging to program, it has been used successfully to achieve speedups of tomographic reconstruction codes such as TxBR [130] or molecular dynamics codes such as NAMD [131]. Emerging standards such as OpenCL (open computing language) will likely help make GPGPU or cGPU (integrated CPU/GPU) even more powerful, lower the cost of use, and increase in popularity. What used to take weeks to complete on a TFLOPS (tera-FLOPS) supercomputer could be accomplished in days on a workstation in the near future. More dedicated application and algorithm specific hardware acceleration for molecular dynamics codes has achieved millisecond scale simulations [132]. Although current force fields have yet to be rigorously tested over such longer time scale simulations, pushing time and length scale limits coupled with experimental validation will help identify critical areas for new methodological development.

Cloud computing refers to the on-demand availability of thousands of processors offered by commercial service providers. Amazon Elastic Cloud 2 (EC2) is offering computation for approximately 10 cents per hour, with enormous impact on virtual screening experiments. Virtual screening services offered as web services are now accessible for interactive use or batch processing of entire library of compounds with transparent access to cloud or cluster resources [133]. The RCS will be available as a computational workflow that will be available to a much wider audience, easily accessible to experimental biologists [134]. Commercial offerings such as the Accelrys Pipeline Pilot tailors to in house chemical compound libraries, whereas the Vision-like environment provides a complete solution to academic researchers in computer aided drug design. Increasingly, both academic and commercial entities will rely on cloud-based virtual screening services that are entirely transparent to the end users. Activities such as the World Community Grid utilize idling computers provides another venue for virtual screening services [135]. From a practical perspective, a brute force approach to cross-dock every resulting structure to every available small molecule compound is counterproductive, regardless of the amount of computational power. Better theories and more efficient computational schemes to allow the selection of conformations most relevant to ligand binding in a predictive manner are still needed. New algorithms such as AutoDock Vina present 10 to 100 fold speed up of AutoDock with equal or better performance [136]. In short, the continued acceleration through algorithmic and

hardware improvement, and cloud-based utility computing will make ensemble based virtual screening accessible to a wider audience, and further its validation and optimization.

CONCLUSIONS

More rigorous molecular dynamics based approaches to include full receptor flexibility are now being applied in larger-scale virtual screening applications, thanks in part to steady increases in computational power and the development of new and improved methods improving computational efficiency of the underlying strategies. The studies highlighted in this review indicate that these ensemble-based methods can yield material discoveries regardless of the choice of force field or MD program. The general correspondence exhibited among the various strategies is a promising feature and indicative of the quality of many of these methods. MD-generated ensemble-based techniques have the potential to provide breakthrough discoveries for many target receptors, not only through the discovery of potential new ligand binding areas, but also through more accurate estimates of free energies of binding for increasingly larger ligand sets.

Acknowledgments

R.E.A. would like to thank Prof. J. Andrew McCammon for his outstanding mentorship and members of the McCammon group for helpful discussions. R.E.A is funded in part by NIH F32-GM077729 and RAC CHE060073. Support from the Howard Hughes Medical Institute, San Diego Supercomputing Center, Accelrys, Inc., the W.M. Keck Foundation, the National Biomedical Computation Resource and the Center for Theoretical Biological Physics is gratefully acknowledged.

References

1. Jain AN. Virtual screening in lead discovery and optimization. Curr Opin Drug Discov Devel. 2004; 7:396–403.
2. 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; 3:935–49. [PubMed: 15520816]
3. Klebe G. Virtual ligand screening: strategies, perspectives and limitations. Drug Discov Today. 2006; 11:580–94. [PubMed: 16793526]
4. Lyne PD. Structure-based virtual screening: an overview. Drug Discov Today. 2002; 7:1047–55. [PubMed: 12546894]
5. Oprea TI, Matter H. Integrating virtual screening in lead discovery. Curr Opin Chem Biol. 2004; 8:349–358. [PubMed: 15288243]
6. Rester U. From virtuality to reality - Virtual screening in lead discovery and lead optimization: a medicinal chemistry perspective. Curr Opin Drug Discov Devel. 2008; 11:559–568.
7. Shoichet BK. Virtual screening of chemical libraries. Nature. 2004; 432:862–865. [PubMed: 15602552]
8. Kontoyianni M, Madhav P, Suchanek E, Seibel W. Theoretical and practical considerations in virtual screening: a beaten field? Curr Med Chem. 2008; 15:107–116. [PubMed: 18220766]
9. Khedkar SA, Malde AK, Coutinho EC. *In Silico* Screening of Ligand Databases: Methods and Applications. Indian J Pharm Sci. 2007; 68:689–696.
10. Ringe D. What makes a binding site a binding site? Curr Opin Struct Biol. 1995; 5:825–829. [PubMed: 8749372]
11. McCammon JA. Theory of biomolecular recognition. Curr Opin Struct Biol. 1998; 8:245–249. [PubMed: 9631300]
12. Gane PJ, Dean PM. Recent advances in structure-based rational drug design. Curr Opin Struct Biol. 2000; 10:401–404. [PubMed: 10981625]
13. Muller-Dethlefs K, Hobza P. Noncovalent interactions: a challenge for experiment and theory. Chem Rev. 2000; 100:143–168. [PubMed: 11749236]

14. Fischer E. Berichte der Deutschen chemischen Gesellschaft. Ber Dt Chem Ges. 1894; 27:2985–2993.
15. Koshland DE. Application of a theory of enzyme specificity to protein synthesis. Proc Natl Acad Sci USA. 1958; 44:98–104. [PubMed: 16590179]
16. Betts MJ, Sternberg MJ. An analysis of conformational changes on protein-protein association: implications for predictive docking. Protein Eng. 1999; 12:271–283. [PubMed: 10325397]
17. Fradera X, De La Cruz X, Silva CH, Gelpi JL, Luque FJ, Orozco M. Ligand-induced changes in the binding sites of proteins. Bioinformatics. 2002; 18:939–348. [PubMed: 12117791]
18. Murray CW, Baxter CA, Frenkel AD. The sensitivity of the results of molecular docking to induced fit effects: application to thrombin, thermolysin and neuraminidase. J Comput Aided Mol Des. 1999; 13:547–562. [PubMed: 10584214]
19. Najmanovich R, Kuttner J, Sobolev V, Edelman M. Side-chain flexibility in proteins upon ligand binding. Proteins. 2000; 39:261–268. [PubMed: 10737948]
20. Paladini DH, Musumeci MA, Carrillo N, Ceccarelli EA. Induced fit and equilibrium dynamics for high catalytic efficiency in ferredoxin-NADP(H) reductases. Biochemistry. 2009; 48:5760–5768. [PubMed: 19435322]
21. Zhao S, Goodsell DS, Olson AJ. Analysis of a data set of paired uncomplexed protein structures: new metrics for side-chain flexibility and model evaluation. Proteins. 2001; 43:271–279. [PubMed: 11288177]
22. Bursavich MG, Rich DH. Designing non-peptide peptidomimetics in the 21st century: inhibitors targeting conformational ensembles. J Med Chem. 2002; 45:541–558. [PubMed: 11806706]
23. Verkhivker GM, Bouzida D, Gehlhaar DK, Rejto PA, Freer ST, Rose PW. Complexity and simplicity of ligand-macromolecule interactions: the energy landscape perspective. Curr Opin Struct Biol. 2002; 12:197–203. [PubMed: 11959497]
24. Onuchic JN, Luthey-Schulten Z, Wolynes PG. Theory of protein folding: the energy landscape perspective. Annu Rev Phys Chem. 1997; 48:545–600. [PubMed: 9348663]
25. Hardin C, Eastwood MP, Prentiss M, Luthey-Schulten Z, Wolynes PG. Folding funnels: the key to robust protein structure prediction. J Comput Chem. 2002; 23:138–46. [PubMed: 11913379]
26. Ma B, Shatsky M, Wolfson HJ, Nussinov R. Multiple diverse ligands binding at a single protein site: a matter of pre-existing populations. Protein Sci. 2002; 11:184–197. [PubMed: 11790828]
27. Chan HS. Kinetics of protein folding. Nature. 1995; 373:664–665. [PubMed: 7854444]
28. Dill KA, Bromberg S, Yue K, Fiebig KM, Yee DP, Thomas PD, Chan HS. Principles of protein folding--a perspective from simple exact models. Protein Sci. 1995; 4:561–602. [PubMed: 7613459]
29. Ma B, Kumar S, Tsai CJ, Nussinov R. Folding funnels and binding mechanisms. Protein Eng. 1999; 12:713–720. [PubMed: 10506280]
30. Knegtel RM, Kuntz ID, Oshiro CM. Molecular docking to ensembles of protein structures. J Mol Biol. 1997; 266:424–440. [PubMed: 9047373]
31. Claussen H, Buning C, Rarey M, Lengauer T. FlexE: efficient molecular docking considering protein structure variations. J Mol Biol. 2001; 308:377–395. [PubMed: 11327774]
32. Osterberg F, Morris GM, Sanner MF, Olson AJ, Goodsell DS. Automated docking to multiple target structures: incorporation of protein mobility and structural water heterogeneity in AutoDock. Proteins. 2002; 46:34–40. [PubMed: 11746701]
33. Barril X, Morley SD. Unveiling the full potential of flexible receptor docking using multiple crystallographic structures. J Med Chem. 2005; 48:4432–4443. [PubMed: 15974595]
34. Bottegoni G, Kufareva I, Totrov M, Abagyan R. Four-dimensional docking: a fast and accurate account of discrete receptor flexibility in ligand docking. J Med Chem. 2009; 52:397–406. [PubMed: 19090659]
35. Huang SY, Zou X. Ensemble docking of multiple protein structures: considering protein structural variations in molecular docking. Proteins. 2007; 66:399–421. [PubMed: 17096427]
36. Rarey M, Kramer B, Lengauer T, Klebe G. A fast flexible docking method using an incremental construction algorithm. J Mol Biol. 1996; 261:470–489. [PubMed: 8780787]

37. Sotriffer CA, Dramburg I. “In situ cross-docking” to simultaneously address multiple targets. J Med Chem. 2005; 48:3122–5. [PubMed: 15857117]
38. Wei BQ, Weaver LH, Ferrari AM, Matthews BW, Shoichet BK. Testing a flexible-receptor docking algorithm in a model binding site. J Mol Biol. 2004; 337:1161–1182. [PubMed: 15046985]
39. Clarage JB, Romo T, Andrews BK, Pettitt BM, Phillips GN Jr. A sampling problem in molecular dynamics simulations of macromolecules. Proc Natl Acad Sci USA. 1995; 92:3288–3292. [PubMed: 7724554]
40. Philippopoulos M, Lim C. Exploring the dynamic information content of a protein NMR structure: comparison of a molecular dynamics simulation with the NMR and X-ray structures of Escherichia coli ribonuclease HI. Proteins. 1999; 36:87–110. [PubMed: 10373009]
41. Salmon L, Bouvignies G, Markwick P, Lakomek N, Showalter S, Li DW, Walter K, Griesinger C, Bruschweiler R, Blackledge M. Protein conformational flexibility from structure-free analysis of NMR dipolar couplings: quantitative and absolute determination of backbone motion in ubiquitin. Angew Chem Int Ed Engl. 2009; 48:4154–4157. [PubMed: 19415702]
42. Ferrari AM, Wei BQ, Costantino L, Shoichet BK. Soft docking and multiple receptor conformations in virtual screening. J Med Chem. 2004; 47:5076–5084. [PubMed: 15456251]
43. Cavasotto CN, Kovacs JA, Abagyan RA. Representing receptor flexibility in ligand docking through relevant normal modes. J Am Chem Soc. 2005; 127:9632–9640. [PubMed: 15984891]
44. Sherman W, Day T, Jacobson MP, Friesner RA, Farid R. Novel procedure for modeling ligand/receptor induced fit effects. J Med Chem. 2006; 49:534–553. [PubMed: 16420040]
45. Teodoro ML, Kavraki LE. Conformational flexibility models for the receptor in structure based drug design. Curr Pharm Des. 2003; 9:1635–1648. [PubMed: 12871062]
46. Totrov M, Abagyan R. Flexible ligand docking to multiple receptor conformations: a practical alternative. Curr Opin Struct Biol. 2008; 18:178–184. [PubMed: 18302984]
47. Hritz J, de Ruiter A, Oostenbrink C. Impact of plasticity and flexibility on docking results for cytochrome P450 2D6: a combined approach of molecular dynamics and ligand docking. J Med Chem. 2008; 51:7469–7477. [PubMed: 18998665]
48. McCammon JA, Gelin BR, Karplus M. Dynamics of folded proteins. Nature. 1977; 267:585–590. [PubMed: 301613]
49. Carlson HA, Masukawa KM, Rubins K, Bushman FD, Jorgensen WL, Lins RD, Briggs JM, McCammon JA. Developing a dynamic pharmacophore model for HIV-1 integrase. J Med Chem. 2000; 43:2100–2114. [PubMed: 10841789]
50. Lins RD, Briggs JM, Straatsma TP, Carlson HA, Greenwald J, Choe S, McCammon JA. Molecular dynamics studies on the HIV-1 integrase catalytic domain. Biophys J. 1999; 76:2999–3011. [PubMed: 10354426]
51. Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. J Amer Chem Soc. 1995; 117:5179–5197.
52. Apra, E.; Windus, TL.; Straatsma, TP.; Bylaska, EJ.; de Jong, W.; Hirata, S.; Valiev, M.; Hackler, M.; Pollack, L.; Kowalski, K.; Harrison, R.; Dupuis, M.; Smith, DMA.; Nieplocha, J.; Tipparaju, V.; Krishnan, M.; Auer, AA.; Brown, E.; Cisneros, G.; Fann, G.; Fruchtl, H.; Garza, J.; Hirao, K.; Kendall, R.; Nichols, J.; Tsemekhman, K.; Wolinski, J.; Anchell, D.; Bernholdt, P.; Borowski, T.; Clark, D.; Clerc, K.; Dachsel, H.; Deegan, M.; Dyall, K.; Elwood, D.; Glendening, E.; Gutowski, M.; Hess, A.; Jaffe, J.; Johnson, B.; Ju, J.; Kobayashi, R.; Kutteh, R.; Lin, Z.; Littlefield, R.; Long, X.; Meng, B.; Nakajima, T.; Niu, S.; Rosing, M.; Sandrone, G.; Stave, M.; Taylor, H.; Thomas, G.; van Lenthe, J.; Wong, A.; Zhang, Z. NWChem, A Computational Chemistry Package for Parallel Computers, Version 5.1. Richland, Washington: 2007. p. 99352-0999.
53. Meagher KL, Carlson HA. Incorporating protein flexibility in structure-based drug discovery: using HIV-1 protease as a test case. J Am Chem Soc. 2004; 126:13276–13281. [PubMed: 15479081]
54. 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; 129:12809–12814. [PubMed: 17902662]

55. Zhong H, Carlson HA. Computational studies and peptidomimetic design for the human p53-MDM2 complex. Proteins. 2005; 58:222–234. [PubMed: 15505803]
56. Lin JH, Perryman AL, Schames JR, McCammon JA. The relaxed complex method: Accommodating receptor flexibility for drug design with an improved scoring scheme. Biopolymers. 2003; 68:47–62. [PubMed: 12579579]
57. Amaro RE, Baron R, McCammon JA. An improved relaxed complex scheme for receptor flexibility in computer-aided drug design. J Comput Aided Mol Des. 2008; 22:693–705. [PubMed: 18196463]
58. Morris GM, Goodsell DS, Halliday RS, Huey R, Hart WE, Belew RK, Olson AJ. Automated docking using a lamarckian genetic algorithm and an empricial binding free energy function. J Comput Chem. 1998; 19:1639–1662.
59. Lin JH, Perryman AL, Schames JR, McCammon JA. Computational drug design accommodating receptor flexibility: the relaxed complex scheme. J Am Chem Soc. 2002; 124:5632–5633. [PubMed: 12010024]
60. Schames JR, Henchman RH, Siegel JS, Sotriffer CA, Ni H, McCammon JA. Discovery of a novel binding trench in HIV integrase. J Med Chem. 2004; 47:1879–1881. [PubMed: 15055986]
61. Hazuda DJ, Anthony NJ, Gomez RP, Jolly SM, Wai JS, Zhuang L, Fisher TE, Embrey M, Guare JP Jr, Egbertson MS, Vacca JP, Huff JR, Felock PJ, Witmer MV, Stillmock KA, Danovich R, Grobler J, Miller MD, Espeseth AS, Jin L, Chen IW, Lin JH, Kassahun K, Ellis JD, Wong BK, Xu W, Pearson PG, Schleif WA, Cortese R, Emini E, Summa V, Holloway MK, Young SD. A naphthyridine carboxamide provides evidence for discordant resistance between mechanistically identical inhibitors of HIV-1 integrase. Proc Natl Acad Sci USA. 2004; 101:11233–11238. [PubMed: 15277684]
62. Summa V, Petrocchi A, Bonelli F, Crescenzi B, Donghi M, Ferrara M, Fiore F, Gardelli C, Gonzalez Paz O, Hazuda DJ, Jones P, Kinzel O, Laufer R, Monteagudo E, Muraglia E, Nizi E, Orvieto F, Pace P, Pescatore G, Scarpelli R, Stillmock K, Witmer MV, Rowley M. Discovery of raltegravir, a potent, selective orally bioavailable HIV-integrase inhibitor for the treatment of HIV-AIDS infection. J Med Chem. 2008; 51:5843–5855. [PubMed: 18763751]
63. Schulten, K.; Phillips, JC.; Kale, L.; Bhatele, A. Biomolecular Modeling in the Era of Petascale Computing. Chapman and Hall/CRC Press: Taylor and Francis Group; New York: 2008.
64. Deng J, Lee KW, Sanchez T, Cui M, Neamati N, Briggs JM. Dynamic receptor-based pharmacophore model development and its application in designing novel HIV-1 integrase inhibitors. J Med Chem. 2005; 48:1496–1505. [PubMed: 15743192]
65. Brigo A, Lee KW, Fogolari F, Mustata GI, Briggs JM. Comparative molecular dynamics simulations of HIV-1 integrase and the T66I/M154I mutant: binding modes and drug resistance to a diketo acid inhibitor. Proteins. 2005; 59:723–741. [PubMed: 15815973]
66. O'Connor SD, Smith PE, al-Obeidi F, Pettitt BM. Quenched molecular dynamics simulations of tuftsin and proposed cyclic analogues. J Med Chem. 1992; 35:2870–2881. [PubMed: 1495018]
67. Wang R, Gao Y, Lai L. LigBuilder: A multi-purpose program for structure-based drug design. J Mol Modelling. 2000; 6:498–516.
68. Accelrys, I. Catalyst 46. San Diego, CA:
69. Cheng LS, Amaro RE, Xu D, Li WW, Arzberger P, McCammon JA. Ensemble-based virtual screening reveals potential novel antiviral compounds for avian influenza neuraminidase. J Med Chem. 2008; 51:3878–3894. [PubMed: 18558668]
70. Amaro RE, Minh DD, Cheng LS, Lindstrom WM Jr, Olson AJ, Lin JH, Li WW, McCammon JA. Remarkable loop flexibility in avian influenza N1 and its implications for antiviral drug design. J Am Chem Soc. 2007; 129:7764–7765. [PubMed: 17539643]
71. Landon MR, Amaro RE, Baron R, Ngan CH, Ozonoff D, McCammon JA, Vajda S. Novel druggable hot spots in avian influenza neuraminidase H5N1 revealed by computational solvent mapping of a reduced and representative receptor ensemble. Chem Biol Drug Des. 2008; 71:106–116. [PubMed: 18205727]
72. Caves LS, Evanseck JD, Karplus M. Locally accessible conformations of proteins: multiple molecular dynamics simulations of crambin. Protein Sci. 1998; 7:649–666. [PubMed: 9541397]

73. Huey R, Morris GM, Olson AJ, Goodsell DS. A semiempirical free energy force field with charge-based desolvation. J Comput Chem. 2007; 28:1145–1152. [PubMed: 17274016]
74. Amaro R, Xiaojin XJ, Xu D, Li WW, Wilson IA, McCammon JA. Novel N1 neuramindase inhibitors discovered through an ensemble-based virtual screen approach. 2009 in preparation.
75. MacKerell AD, Bashford D, Bellott M, Dunbrack RL, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WE, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiorkiewicz-Kuczera J, Yin D, Karplus M. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. J Phys Chem B. 1998; 102:3586–3616.
76. Mackerell AD Jr. Empirical force fields for biological macromolecules: overview and issues. J Comput Chem. 2004; 25:1584–1604. [PubMed: 15264253]
77. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, et al. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. J Comp Chem. 1983; 4:187–217.
78. Zhong S, Chen X, Zhu X, Dziegielewska B, Bachman KE, Ellenberger T, Ballin JD, Wilson GM, Tomkinson AE, MacKerell AD Jr. Identification and validation of human DNA ligase inhibitors using computer-aided drug design. J Med Chem. 2008; 51:4553–4562. [PubMed: 18630893]
79. Ewing TJ, Makino S, Skillman AG, Kuntz ID. DOCK 4.0: search strategies for automated docking of flexible molecule databases. J Comput Aided Mol Des. 2001; 15:411–428. [PubMed: 11394736]
80. O'Donoghue P, Luthey-Schulten Z. Evolutionary profiles derived from the QR factorization of multiple structural alignments gives an economy of information. J Mol Biol. 2005; 346:875–894. [PubMed: 15713469]
81. Amaro RE, Schnaufer A, Interthal H, Hol W, Stuart KD, McCammon JA. Discovery of drug-like inhibitors of an essential RNA-editing ligase in Trypanosoma brucei. Proc Natl Acad Sci USA. 2008; 105:17278–17283. [PubMed: 18981420]
82. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kale L, Schulten K. Scalable molecular dynamics with NAMD. J Comput Chem. 2005; 26:1781–1802. [PubMed: 16222654]
83. Amaro R, Swift R, McCammon JA. Functional and structural insights revealed by molecular dynamics simulations of an essential Rna editing ligase in *trypanosoma brucei*. PLoS Negl Trop Dis. 2007; 1(2):e68. [PubMed: 18060084]
84. Barillari C, Taylor J, Viner R, Essex JW. Classification of water molecules in protein binding sites. J Am Chem Soc. 2007; 129:2577–2587. [PubMed: 17288418]
85. Lu Y, Wang R, Yang CY, Wang S. Analysis of ligand-bound water molecules in high-resolution crystal structures of protein-ligand complexes. J Chem Inf Model. 2007; 47:668–675. [PubMed: 17266298]
86. Hamelberg, D.; McCammon, JA. Dealing with bound waters in a site: Do they leave or do they stay?. In: Stroud, RM., editor. Computational and Structural Approaches to Drug Discovery. Royal Society of Chemistry; 2007. p. 95-110.
87. Huang N, Shoichet BK. Exploiting ordered waters in molecular docking. J Med Chem. 2008; 51:4862–4865. [PubMed: 18680357]
88. Denisov VP, Schlessman JL, Garcia-Moreno EB, Halle B. Stabilization of internal charges in a protein: water penetration or conformational change? Biophys J. 2004; 87:3982–3994. [PubMed: 15377517]
89. Hamelberg D, McCammon JA. Standard free energy of releasing a localized water molecule from the binding pockets of proteins: double-decoupling method. J Am Chem Soc. 2004; 126:7683–7689. [PubMed: 15198616]
90. Hamelberg D, Mongan J, McCammon JA. Accelerated molecular dynamics: a promising and efficient simulation method for biomolecules. J Chem Phys. 2004; 120:11919–11929. [PubMed: 15268227]
91. Hamelberg D, de Oliveira CA, McCammon JA. Sampling of slow diffusive conformational transitions with accelerated molecular dynamics. J Chem Phys. 2007; 127:155102. [PubMed: 17949218]

92. Minh DD, Hamelberg D, McCammon JA. Accelerated entropy estimates with accelerated dynamics. J Chem Phys. 2007; 127:154105. [PubMed: 17949130]

93. Markwick PR, Bouvignies G, Blackledge M. Exploring multiple timescale motions in protein GB3 using accelerated molecular dynamics and NMR spectroscopy. J Am Chem Soc. 2007; 129:4724–4730. [PubMed: 17375925]

94. Bashford D, Case DA. Generalized born models of macromolecular solvation effects. Annu Rev Phys Chem. 2000; 51:129–152. [PubMed: 11031278]

95. Simmerling C, Strockbine B, Roitberg AE. All-atom structure prediction and folding simulations of a stable protein. J Am Chem Soc. 2002; 124:11258–11259. [PubMed: 12236726]

96. Onufriev A, Bashford D, Case DA. Exploring protein native states and large-scale conformational changes with a modified generalized born model. Proteins. 2004; 55:383–94. [PubMed: 15048829]

97. Tsui V, Case DA. Molecular dynamics simulations of nucleic acids with a generalized born solvation model. J Am Chem Soc. 2000; 122:2489–2498.

98. Amaro RE, Cheng X, Ivanov I, Xu D, McCammon JA. Characterizing loop dynamics and ligand recognition in human- and avian-type influenza neuraminidases via generalized born molecular dynamics and end-point free energy calculations. J Am Chem Soc. 2009; 131:4702–4709. [PubMed: 19296611]

99. Ruscio JZ, Onufriev A. A Computational Study of nucleosomal DNA flexibility. Biophys J. 2006; 91:4121–4132. [PubMed: 16891359]

100. Fan H, Mark AE, Zhu J, Honig B. Comparative study of generalized Born models: protein dynamics. Proc Natl Acad Sci USA. 2005; 102:6760–6764. [PubMed: 15814616]

101. Corcho FJ, Mokoena P, Bisetty K, Perez JJ. Molecular dynamics (MD) simulations of VIP and PACAP27. Biopolymers. 2009; 91:391–400. [PubMed: 19156830]

102. Lybrand TP, McCammon JA, Wipff G. Theoretical calculation of relative binding affinity in host-guest systems. Proc Natl Acad Sci USA. 1986; 83:833–835. [PubMed: 3456569]

103. Gilson MK, Given JA, Bush BL, McCammon JA. The statistical-thermodynamic basis for computation of binding affinities: a critical review. Biophys J. 1997; 72:1047–1069. [PubMed: 9138555]

104. Aqvist J, Luzhkov VB, Brandsdal BO. Ligand binding affinities from MD simulations. Acc Chem Res. 2002; 35:358–365. [PubMed: 12069620]

105. Gohlke H, Klebe G. Approaches to the description and prediction of the binding affinity of small-molecule ligands to macromolecular receptors. Angew Chem Int Ed Engl. 2002; 41:2644–2676. [PubMed: 12203463]

106. Jorgensen WL. The many roles of computation in drug discovery. Science. 2004; 303:1813–8. [PubMed: 15031495]

107. Adcock SA, McCammon JA. Molecular dynamics: survey of methods for simulating the activity of proteins. Chem Rev. 2006; 106:1589–1615. [PubMed: 16683746]

108. Tembe BL, McCammon JA. Ligand-receptor interactions. Comput Chem. 1984; 8:281–283.

109. Wong CF, McCammon JA. Dynamics and design of enzymes and inhibitors. J Amer Chem Soc. 1986; 108:3830–3832.

110. King, PM. Free energy via molecular simulation: a primer. Vol. 2. Escom; Leiden: 1993. p. 267-314.

111. Beveridge DL, DiCapua FM. Free energy via molecular simulation: applications to chemical and biomolecular systems. Annu Rev Biophys Biophys Chem. 1989; 18:431–492. [PubMed: 2660832]

112. Straatsma TP, McCammon JA. Theoretical calculations of relative affinities of binding. Methods Enzymol. 1991; 202:497–511. [PubMed: 1784186]

113. Lamb ML, Jorgensen WL. Computational approaches to molecular recognition. Curr Opin Chem Biol. 1997; 1:449–457. [PubMed: 9667895]

114. Chipot C, Rozanska X, Dixit SB. Can free energy calculations be fast and accurate at the same time? Binding of low-affinity, non-peptide inhibitors to the SH2 domain of the src protein. J Comput Aided Mol Des. 2005; 19:765–770. [PubMed: 16365699]

115. Reinhardt WP, Miller MA, Amon LM. Why is it so difficult to simulate entropies, free energies, and their differences? Acc Chem Res. 2001; 34:607–614. [PubMed: 11456478]

116. van Gunsteren WF, Bakowies D, Baron R, Chandrasekhar I, Christen M, Daura X, Gee P, Geerke DP, Glattli A, Hunenberger PH, Kastenholz MA, Oostenbrink C, Schenk M, Trzesniak D, van der Vegt NF, Yu HB. Biomolecular modeling: Goals, problems, perspectives. Angew Chem Int Ed Engl. 2006; 45:4064–92. [PubMed: 16761306]

117. van Gunsteren WF, Daura X, Mark AE. Computation of free energy. Helv Chim Acta. 2002; 85:3113–3129.

118. van Gunsteren, WF.; Beutler, TC.; Fraternali, F.; King, PM.; Mark, AE.; Smith, PE. Computation of free energy in practice: choice of approximations and accuracy limiting factors. Vol. 2. Escom; Leiden: 1993. p. 315-348.

119. Hunenberger PH, Helms V, Narayana N, Taylor SS, McCammon JA. Determinants of ligand binding to cAMP-dependent protein kinase. Biochemistry. 1999; 38:2358–2366. [PubMed: 10029529]

120. Steinbrecher T, Case DA, Labahn A. A multistep approach to structure-based drug design: studying ligand binding at the human neutrophil elastase. J Med Chem. 2006; 49:1837–1844. [PubMed: 16539369]

121. Bowers, KJ.; Chow, E.; Xu, H.; Dror, RO.; Eastwood, MP.; Gregersen, BA.; Klepeis, JL.; Kolossváry, I.; Moraes, MA.; Sacerdoti, FD.; Salmon, JK.; Shan, Y.; Shaw, DE. ACM. Scalable Algorithms for Molecular Dynamics Simulations on Commodity Clusters. Proceedings of the ACM/IEEE Conference on Supercomputing (SC06); Tampa, Florida. 2006.

122. Kollman PA, Massova I, Reyes C, Kuhn B, Huo S, Chong L, Lee M, Duan Y, Wang W, Donini O, Cieplak P, Srinivasan J, Case DA, Cheatham TE. Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. Acc Chem Res. 2000; 33:889–897. [PubMed: 11123888]

123. Massova I, Kollman PA. Combined molecular mechanical and continuum solvent approach (MM-PBSA/GBSA) to predict ligand binding. Perspect Drug Discov Des. 2000; 18:113–135.

124. Almlof M, Brandsdal BO, Aqvist J. Binding affinity prediction with different force fields: Examination of the linear interaction energy method. J Comput Chem. 2004; 25:1242–1254. [PubMed: 15139037]

125. Aqvist J, Medina C, Samuelsson JE. A new method for predicting binding affinity in computer-aided drug design. Protein Eng. 1994; 7:385–391. [PubMed: 8177887]

126. Liu H, Mark AE, van Gunsteren WF. Estimating the relative free energy of different molecular states with respect to a single reference state. J Phys Chem. 1996; 100:9485–9494.

127. Mark AE, Xu Y, Liu H, van Gunsteren WF. Rapid non-empirical approaches for estimating relative binding free energies. Acta Biochim Pol. 1995; 42:525–535. [PubMed: 8852343]

128. Oostenbrink C, van Gunsteren WF. Single-step perturbations to calculate free energy differences from unphysical reference states: limits on size, flexibility, and character. J Comput Chem. 2003; 24:1730–1739. [PubMed: 12964191]

129. Zagrovic B, van Gunsteren WF. Computational analysis of the mechanism and thermodynamics of inhibition of phosphodiesterase 5A by synthetic ligands. J Chem Theory Comput. 2007; 3:301–311.

130. Lawrence, A.; Phan, S.; Singh, R. Parallel processing and large-field electron microscope tomography, CSIE 2009. IEEE Computer Society; Los Angeles/Anaheim: 2009. In Press

131. Stone JE, Phillips JC, Freddolino PL, Hardy DJ, Trabuco LG, Schulten K. Accelerating molecular modeling applications with graphics processors. J Comput Chem. 2007; 28:2618–40. [PubMed: 17894371]

132. Dror, RO.; Salmon, JK.; Grossman, JP.; Mackenzie, KM.; Bank, JA.; Young, C.; Deneroff, MM.; Batson, B.; Bowers, KJ.; Chow, E.; Eastwood, MP.; Ierardi, DJ.; Klepels, JL.; Kurskin, JS.; Larson, RH.; Lindorff-Larsen, K.; Maragakis, P.; Moraes, MA.; Plana, S.; Shan, Y.; Towles, B.; Shaw, DE. Supercomputing 2009. Portland, OR: 2009. Millisecond-Scale Molecular Dynamics Simulation on Anton.

133. Krishnan, S.; Clementi, L.; Ren, J.; Papadopoulos, P.; Li, WW. Design and evaluation of Opal 2: toolkit for scientific software as a service; International Workshop on Cloud Services, International Conference on Web Services 2009; Los Angeles,. 2009. In Press

134. Clementi, L.; Krishnan, S.; Goodman, W.; Ren, J.; Li, WW.; Arzberger, PW.; Vareille, G.; Sanner, M. Services oriented architecture for managing workflows of avian flu grid; IEEE eScience 2008; Indianapolis, IN. 2008.

135. Chang MW, Lindstrom W, Olson AJ, Belew RK. Analysis of HIV wild-type and mutant structures *via in silico* docking against diverse ligand libraries. J Chem Inf Model. 2007; 47:1258–1262. [PubMed: 17447753]

136. Trott O, Olson AJ. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem. 2009 [Epub ahead of print].

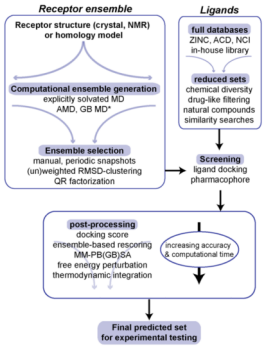

Fig. 1.

General workflow for ensemble-based virtual screen experiment. Blue arrows indicate size of data sets (i.e. increasing or decreasing) at each step; * denotes emerging methods that have not yet been tested. (AMD: accelerated molecular dynamics, GB MD: generalized Born molecular dynamics, RMSD: root-mean-square-deviation, ZINC – ZINC Is Not Commerical, ACD: Available Chemical Database, NCI: National Cancer Institute, MM-PB(GB)SA: Molecular Mechanics – Poisson-Boltzmann (Generalized Born) Surface Area).

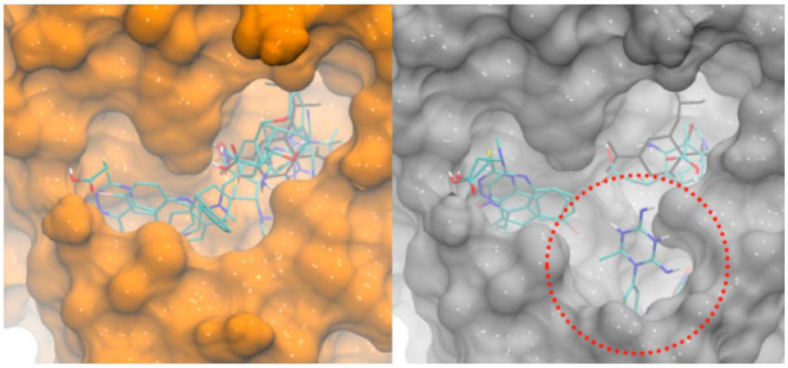

Fig. 2.

MD generated conformational ensemble used for virtual screening of the flexible avian influenza neuraminidase receptor (PDB 2HU4). Closed binding pocket (orange, left panel) from crystal structures; wide open 150- and 430-loop areas (grey, right panel) from MD simulations [69]. Select compounds from the virtual screen are shown docked to these areas; oseltamivir shown in dark gray (open pocket highlighted in red circle).

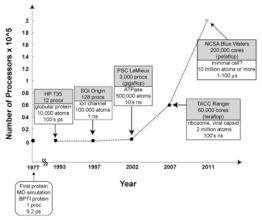

Fig. 3.

Evolution of publicly available compute power since 1993. Hardware advances (machine name, number of processors) are shown in grey, whereas system advances (example system, number of atoms, and average timescale for simulations) are shown in white.