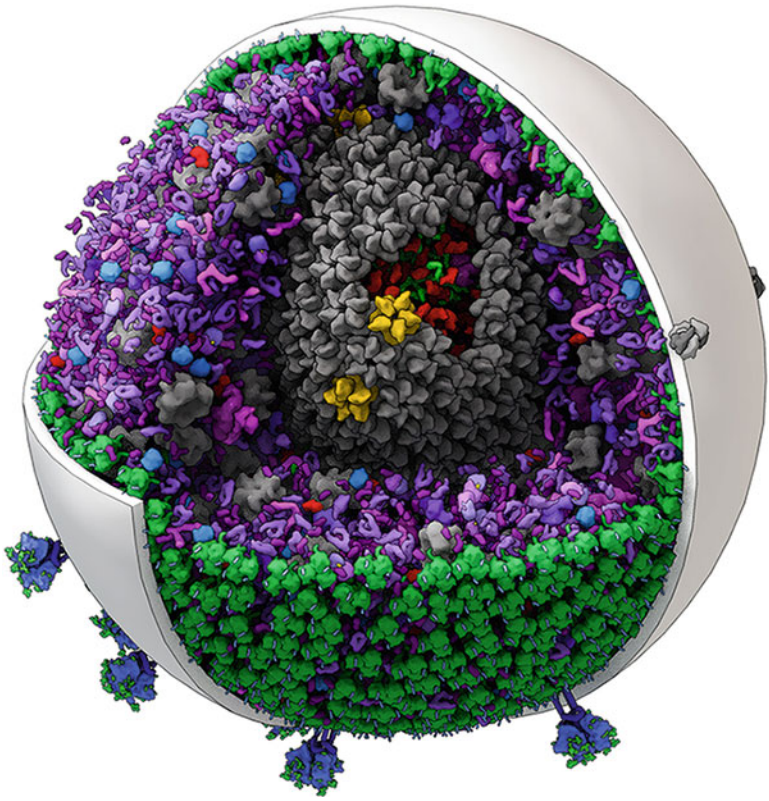
# Computational Challenges of Structure-Based Approaches Applied to HIV

**Stefano Forli and Arthur J. Olson**



Three-dimensional model of mature HIV. A cutaway view of mature HIV includes the capsid (*gray*, with pentamers in *yellow*) and nucleocapsid (*red*), matrix protein (*green*), accessory proteins (*magenta*), membrane (*white*), and envelope protein (*blue*). The model was generated using CellPACK by Graham Johnson

S. Forli · A.J. Olson (✉)
MGL, Department of Integrative Structural and Computational
Biology and HIV Interaction and Viral Evolution Center,
The Scripps Research Institute, 10550 N. Torrey Pines Rd., La Jolla, CA 92037, USA
e-mail: olson@scripps.edu



Current Topics in Microbiology and Immunology (2015) 389:31–51
DOI 10.1007/82_2015_432
Published Online: 25 February 2015

**Abstract** Here, we review some of the opportunities and challenges that we face in computational modeling of HIV therapeutic targets and structural biology, both in terms of methodology development and structure-based drug design (SBDD). Computational methods have provided fundamental support to HIV research since the initial structural studies, helping to unravel details of HIV biology. Computational models have proved to be a powerful tool to analyze and understand the impact of mutations and to overcome their structural and functional influence in drug resistance. With the availability of structural data, in silico experiments have been instrumental in exploiting and improving interactions between drugs and viral targets, such as HIV protease, reverse transcriptase, and integrase. Issues such as viral target dynamics and mutational variability, as well as the role of water and estimates of binding free energy in characterizing ligand interactions, are areas of active computational research. Ever-increasing computational resources and theoretical and algorithmic advances have played a significant role in progress to date, and we envision a continually expanding role for computational methods in our understanding of HIV biology and SBDD in the future.

**Abbreviations**

| | |
|---|---|
| BEDAM | Binding energy distribution analysis method |
| CCD | Catalytic core domain |
| BSI | Backscattering interferometry |
| DSF | Differential scanning fluorimetry |
| DTP | Developmental Therapeutics Program |
| FA@H | FightAids@Home |
| FBDD | Fragment-based drug design |
| HAART | Highly active antiretroviral therapy |
| HIV | Human immunodeficiency virus |
| HTVS | High-throughput virtual screening |
| IN | Integrase |
| INSTI | IN strand transfer inhibitor |
| LEGDF | Lens epithelium-derived growth factor |
| MD | Molecular dynamics |
| MW | Molecular weight |
| NMA | Normal mode analysis |
| PDB | Protein Data Bank |
| PPI | Protein–protein interaction |
| PR | Protease |
| RC | Relaxed complex |
| RH | RNase H |
| RT | Reverse transcriptase |
| SAMPL | Statistical Assessment of Modeling of Proteins and Ligands |
| SBDD | Structure-based drug design |
| WCG | World Community Grid |

## Contents



## 1 Introduction

The AIDS epidemic and structure-based drug discovery have evolved over the same historical time period, with each having significant impact on the other. In the early 1980s, AIDS was first recognized as a new epidemic and a serious threat to human health. During that same time period, protein crystallography was coming into its own, with new structures appearing on an ever-increasing basis. In the mid-1980s, the discovery that the causative agent of AIDS was HIV (then called HTLV III) enabled researchers to culture the virus, which set the stage for producing viral proteins in sufficient quantities to undertake crystallographic studies. During that same time period, computational chemists were starting to utilize protein crystal structures to try to discover new compounds to inhibit protein function (Goodsell and Olson 1990; Kuntz et al. 1982; Moon and Howe 1991; Marshall 1987). A major computational development at that time was the development of automated docking programs that could predict how a chemical compound could bind to the active site of an enzyme, propelling the field of rational structure-based drug design (SBDD). By 1990, the crystal structure of HIV protease had been determined, and efforts were underway to develop active site inhibitors based on what was known about its structure and function. Also at that time, the National Institutes of Health made a significant push to understand HIV protein structures, and the National Institute for General Medical Sciences funded its Structural Biology of HIV/AIDS Program, under the leadership of Dr. Marvin Cassman. By the mid-1990s, the first structure-based HIV protease inhibitors were approved for the treatment of AIDS, enabling the development of highly active antiretroviral therapy (HAART), which in turn resulted in a rapid decline of AIDS deaths where such treatment was available.

Over the intervening years, thanks in part to the NIGMS Structure Biology of HIV/AIDS Program, HIV has become one of the most structurally well characterized of human viruses and is a model system for understanding the complex relationships between virus and host from a structural perspective. Concomitant with these developments, computational capability has evolved with the exponential power of Moore's Law along with theoretical and algorithmic advances

encoded in computer programs. The problems that can now be approached go well beyond the early drug design paradigm involving a single ligand docked to a single, static crystal structure of the target. The more we explore the evolution and mechanisms of drug resistance, for instance, the more complex the computational modeling needs to be. Here, we review some of the challenges that we face, and the directions that we and others are developing.

## 2 Computational Structure-Based Drug Design

New findings constantly provide alternative targets or alternative locations on known targets. Also, HIV biological targets have been a perfect test bed for the design and validation of computational methods. Despite the fact that the virus encodes only 15 mature proteins (Engelman and Cherepanov 2012), it exploits many host proteins during its entire life cycle, from the initial interaction with the cell membrane, during integration and replication process, until maturation and assembly. For this reason, the activity of a single protein such as integrase (IN) can be inhibited either by the classical inhibition mechanism via targeting its active site (Hare et al. 2010) (e.g., protease and integrase inhibitors) and interfering directly with the DNA strand transfer activity or by inhibiting its interaction with p75/LEDGF via protein–protein interaction (PPI) inhibition. (Christ et al. 2010) Our laboratory has been involved for more than two decades in active research and computational modeling of HIV therapeutic targets, both in methodology development and structure-driven drug design. Since the wide availability of experimental data has made HIV an ideal system for development of new methodologies, our application of SBDD to HIV has enabled us to analyze most of the common issues and limitations related to molecular modeling methodology.

## 3 Structure-Based Drug Design

HIV PR is often cited as an example of the successful application of SBDD, and it has been a test bed for most of the techniques used currently in SBDD. The first PR structure was solved in 1989 (Navia et al. 1989) but before that, accurate models had been generated using homology modeling (Weber 1990). Since the time of the first HIV PR crystal structure deposition in the Protein Data Bank (PDB) (Navia et al. 1989), hundreds of more PR structures have followed. The ready availability of structural information supported the early SBDD research and was instrumental in the design of the first PR inhibitors. In fact, saquinavir, the first protease inhibitor, was approved by the FDA in 1995, only a few years after the availability of the first PR structure. Its availability leads to a dramatic reduction of AIDS mortality (Hall et al. 2008). Modeling was extensively applied in the rational design based on the vast amount of structural data gathered for PR, supporting exploratory

and advanced approaches such as in situ click chemistry (Whiting et al. 2006). Ligand–protein docking is among the most successful techniques applied in SBDD (Cosconati et al. 2010) The main issues to be addressed in molecular docking are prediction of the correct ligand configuration in the target protein (i.e., global search algorithm) and proper energetic scoring of the pose (i.e., scoring function). Accuracy in predicting the ligand pose is dependent on both ligand structural complexity and receptor topography. In particular, PR inhibitors are relatively large and flexible molecules with MW exceeding 650 dalton and with numerous freely rotatable bonds. Docking very flexible ligands has presented a significant search issue, especially in the early days of modeling due to the considerable computational power required to efficiently sample the conformational space during a global search (Fig. 1). Poses generated during the global search are ranked using a scoring function that is usually derived from an empirically tuned physics-based energy model (Halperin et al. 2002; Warren et al. 2006; Wang et al. 2003). Most scoring functions contain simplified descriptions of energetic terms (e.g., electrostatic) trading accuracy and precision for speed and efficiency. In fact, a small computational footprint is essential in docking, where the scoring function ("energy evaluation") is called at every step of the extensive search procedure. Therefore, empirical calibration of each scoring function term is essential to build a robust model that is able to describe ligand–receptor interactions. A considerable number of complexes of PR with inhibitors were used in the training and validation of the force field of our docking software AutoDock (Morris et al. 1998). Over time, advances in hardware capability and software search algorithms have been reflected in marked improvement in docking performance. Recently, we developed a new version of AutoDock, called AutoDock Vina (Trott and Olson 2010), to exploit advantages provided by multi-core processor architecture. The improved software design in Vina includes a parallelized search algorithm that enables handling extremely flexible ligands and speed calculation by about two orders of magnitude (Trott and Olson 2010).

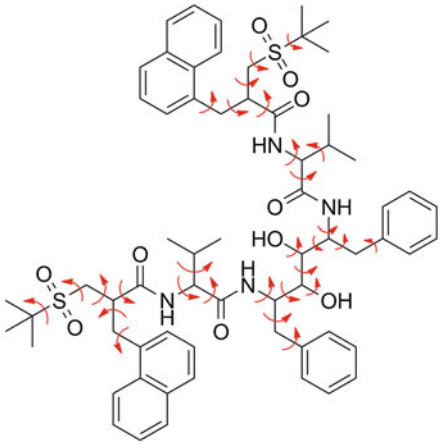



**Fig. 1** Scheme of 29 rotatable bonds that are be sampled during docking of a very potent (0.3 nM) PR active site inhibitor (PDB: *1vij*) (Lange-Savage et al. 1997)

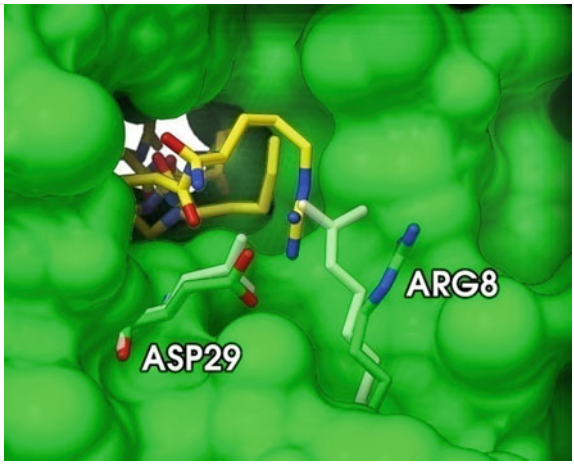



**Fig. 2** HIV-1 protease side chain movements upon active site engagement: unbound state (side chains: *white sticks*; PDB: *4ej8*) (Perryman et al. 2010b); bound state (in complex with a substrate peptide corresponding to reverse transcriptase (RT) and RNase H (RH) domains, derived from the *pol* polypeptide: *yellow sticks*; side chains: *green sticks*; PDB: *4ep2*) (Alvizo et al. 2012)

One of the major limitations of most docking programs, including AutoDock, is the fact that the target protein is treated as rigid during the entire calculation. This becomes particularly important when receptor flexibility plays an important role, as in the case of PR. For example, Arg8–Asp29 salt bridges established between the two monomers act as gatekeepers at both ends of the active site tunnel (Fig. 2). Rigid receptor docking cannot represent this salt bridge opening and prevents the binding of large ligands (Österberg et al. 2002). Version 4 of AutoDock added support for selected flexible receptor side chains, and it was validated on a cross-docking experiment with 87 PR complexes demonstrating a significant improvement in performance (Morris et al. 2009). Results showed that allowing side chains to accommodate to ligand interaction during docking increased both accuracy and rate of success. Another important issue in docking is the role of water molecules near the active site, influencing ligand binding and its interaction energy. In PR, a water molecule (often referred to as W301 (Kiso et al. 1999), Fig. 3) plays a structural role in the active site. W301 was extensively investigated (Nicholson et al. 1995; Suresh et al. 2008) and was found to be stably bound to the protein structure, adapting the active site topography to the substrate, while mediating its interactions with the flaps, mainly with residues Ile50/150. Other waters beside W301 have also been reported mediating secondary ligand–receptor interactions (Baldwin et al. 1995; Wang et al. 1996) (Fig. 3). Both interaction mediation and displacement of W301 have been exploited in PR drug design. In particular, cyclic urea derivatives (Grzesiek et al. 1994; Hodge et al. 1996; Ala et al. 1998; Jadhav et al. 1997) were designed to bind while displacing W301.

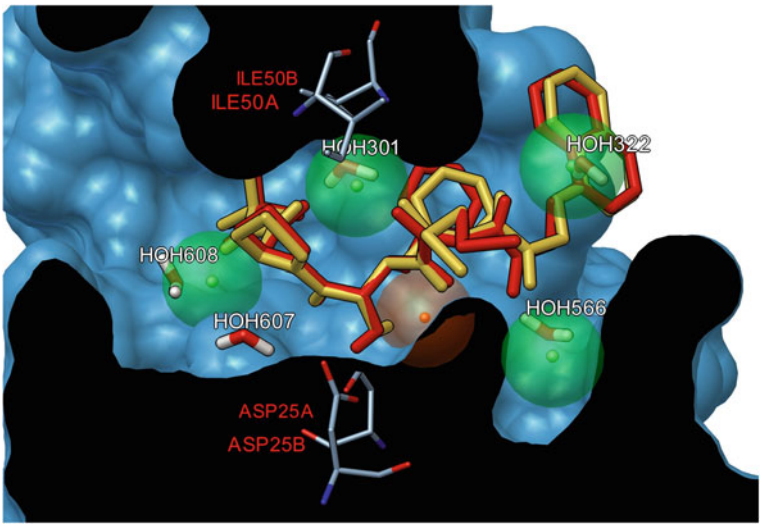



**Fig. 3** Docking with water prediction (Forli and Olson 2012) of HIV-1 protease inhibitors (PDB: *2zye*) (Adachi et al. 2009). Ligand is shown as *thick* sticks (*red* experimental, *yellow* docked). Waters are shown as sticks (experimental) or spheres (predicted). Spheres are colored by prediction accuracy (*green* high, *red* low)

Typically, waters are removed from the target protein structure prior to docking, with the exception of those known to be conserved in the binding site. However, this implies that different target structures, hydrated and dehydrated forms, need to be used when hydration states are unknown or variable. For this reason, we developed a protocol to dynamically predict the presence and the role of water molecules in binding sites using HIV PR complexes to validate the method (Forli and Olson 2012). The method is based on the prediction of potential hydration sites on the ligand, and these predictions are then used to estimate the role of each possible water separately (Fig. 3). This protocol significantly improved docking performance for cases where waters mediate ligand binding, especially in the case of small chemical fragment docking. In fact, it has become clear that individual waters play a large role in binding of molecular fragments (e.g., MW < 150D). The prediction of such waters in docking is thus critical to reproduce the correct binding mode for these systems.

Accurate estimations of desolvation energy, and changes in flexibility profiles of both protein and ligand are required in order to achieve the computational ability to discriminate between binders and non-binders. Although most scoring functions used in molecular docking provide a fast approximation for energy estimation mainly based on the enthalpic components (i.e., van der Waals, electrostatic and hydrogen bonding interactions), they usually have a very approximate accounting of the entropic effects and therefore provide a poor representation of the binding thermodynamics. For example, the subtle balance between enthalpy and entropy components in free energy of binding has been observed to be extremely important

for PR ligand binding (Lu and Li 2010; King et al. 2012). An efficient way of dealing with the need for accuracy would be to refine promising docking results by using more sophisticated and computationally intensive methods. Such protocols can enable a reduction in the number of false-positive results (i.e., non-binder ligands predicted to bind) maximizing efficiency in the subsequent selection, acquisition, and testing of chemical compounds in biological assays. We successfully demonstrated this pipeline in the fourth edition of the Statistical Assessment of Modeling of Proteins and Ligands (SAMPL4) (Mobley et al. 2014) in collaboration with Ron Levy at Temple University (Gallicchio et al. 2014; Perryman et al. 2014). The goal of the SAMPL4 challenge was a blind prediction of the binding of inhibitors into IN catalytic core domain (CCD). Essentially, our pipeline comprised two steps: (1) a docking phase using AutoDock Vina coupled with a common pharmacophore model (Perryman et al. 2014) and (2) a series of molecular dynamics (MD) simulations performed using the binding energy distribution analysis method (BEDAM) (Gallicchio et al. 2010, 2014). BEDAM is a Hamiltonian replica exchange free energy simulation that provides physics-based binding free energy model capable of high accuracy in predicting reorganizational energetics and desolvation and entropic changes of receptor and ligand upon binding. The application of this pipeline dramatically reduced the number of false-positive results, and the measured enrichment factor achieved ranked best among computational submissions in the challenge.

Another consideration from the analysis of the SAMPL4 results was that the application of available knowledge-based structural information allows improved performance over strictly force-field-based approaches. In fact, the overall best entry in the challenge consisted of manually identified results by an expert in the specific system, while our computational method using pharmacophore descriptors based upon known structures provided the best automated results. The pharmacophore model was built using structures of known IN binders aligned in their respective binding sites (Perryman et al. 2014). A newly designed pharmacophoric engine was then used to extract common features (e.g., hydrogen bond acceptor/donor, and aromatic rings) recapitulating the interaction pattern of each ligand set. The pharmacophoric models were then used to re-score dockings, significantly increasing enrichment factor over docking results alone.

These results emphasize that the availability of high-quality models of biological target structures is critical in performing SBDD. HIV represents an ideal case, due to the limited number of viral protein components, with the great majority already structurally resolved. What is also essential, however, is the prediction of druggable sites on the protein structure where potential inhibitors can bind. While enzymes typically have easily identifiable active sites, structural proteins typically do not present obviously targetable binding regions. Moreover, even in cases where a binding site is already known, the identification of alternative binding locations (e.g., allosteric sites) can provide new druggable targets to exploit in the search for strategies to defeat the evolution of drug resistance. Several programs (Liang et al. 2006; Laurie and Jackson 2005; Capra et al. 2009) and Web servers (Volkamer et al. 2012; Binkowski et al. 2003) using a variety of methods are available to

perform calculations to identify such sites on protein structures. Among these is AutoLigand, a tool developed in our laboratory. AutoLigand scans topographical and chemical features in the input structure to predict the location of high-affinity regions where a candidate drug molecule can bind and interact effectively (Harris et al. 2008). AutoLigand is derived from the AutoDock force field, to characterize the optimal size, shape, and chemical characteristics of a ligand for a promising site. The data set of 280 diverse ligand–protein complexes on which the software was validated and tested included several PR structures. Validation results showed very good results (73 % success rate in identification of both the correct location of the known binding sites and the volume of the known binder; for apo structures, 80 % success rate in identifying correct site location). Interestingly, the optimal AutoLigand envelope predicted for PR active site precisely matched the experimental structure of the high-affinity inhibitor in the complex. AutoLigand has been instrumental in the identification of one of the putative PR allosteric sites (Perryman et al. 2010b); calculations performed on this exosite were in good agreement with the structure of the methyl-cyclohexanol fragment 4D9 identified through crystallographic screenings (Figs. 4 and 5).

Another way to inhibit viral functions is to interfere with interactions between viral proteins, or between viral and host proteins (i.e., PPI inhibition). This approach is more technically challenging, mainly due to the characteristics of

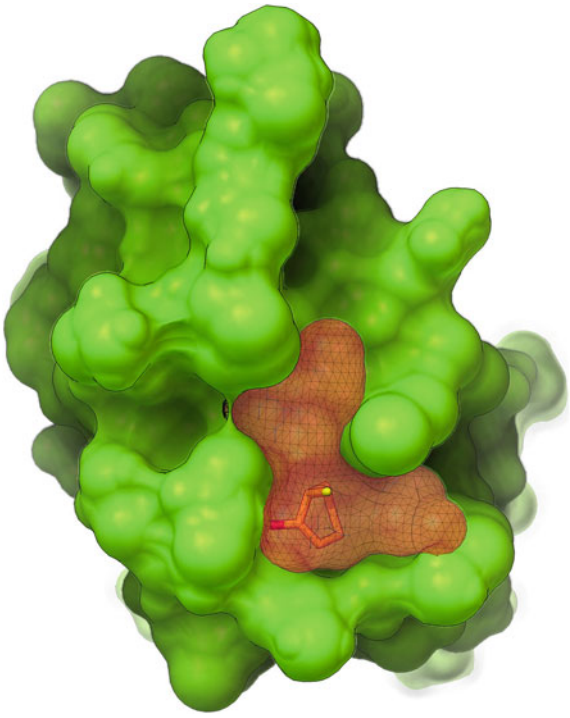


**Fig. 4** AutoLigand result on HIV-1 protease showing the predicted location of the exosite (Perryman et al. 2010b). Protein surface is shown as MSMS surface (Sanner et al. 1996) (*green*); crystal structure of the 4D9 fragment (PDB: *3kf0*) bound in the exosite is shown as sticks (*yellow*); AutoLigand fill volume is shown as mesh (*red*)

structural interfaces involved. Protein–protein interfaces, in fact, tend to be larger and relatively flat, when compared to well-defined cavities of enzyme catalytic sites, for example. Nevertheless, this approach has led to success, such as the recent results from Christ and co-workers that identified a novel inhibitor of the interaction between IN and the human protein LEDGF/p75 (Christ et al. 2010) which aids in HIV provirus integration into human chromatin. The inhibition of this interaction provides an alternative mechanism to interfere with IN activity in infected cells. The main advantage of interrupting these host–virus interactions for HIV therapy is the restricted mutational range that the virus can sample, being forced to maintain interactions with highly conserved human proteins, but resistance mutations are already known (Christ et al. 2010). We are applying several modeling techniques (docking, pharmacophoric analysis, and MD) to design better ligands in the IN/LEDGF site to overcome the development of resistance.

## 4 Large-Scale Modeling

Along with other biological targets, the study of HIV has benefited from the constant increase in computational efficiency. With the availability of ever-growing computer resources and improved software algorithms, it has been possible to increase the scope of computational analyses by orders of magnitude. The most representative example of such large computational experiments is high-throughput virtual screening (HTVS) campaigns, where up to millions of chemical compounds are docked against one or more target structures. The most significant advantage of screening of large compound libraries is the dramatic expansion of the chemical space explored, where new hits and potential drugs can be identified. However, even with this computational power, different ways to reduce the dimensionality of the large-scale dockings are needed to more efficiently sample chemical space (estimated at $10^{60}$ compounds). A widely used approach is the fragment-based drug design (FBDD) approach (Hajduk and Greer 2007), where libraries of small chemical fragments are used instead of drug-like compounds to identify new hits. We successfully applied FBDD to crystallographically screen a small fragment library of 378 compounds on PR (Perryman et al. 2010b), identifying the first three non-active site binders of PR (4D9 in the exosite, 1F1 and 2F4 in the flap site, Fig. 5). These fragments provided indications about chemical groups likely to interact with several potential allosteric pockets on PR. Structural information gathered with fragments exploited to build a pharmacophore model using our Common Pharmacophore Engine (Perryman et al. 2014). The model enabled filtering and identification of larger molecules that possess similar chemical characteristics of the fragments. These molecules have been crystallized in the predicted allosteric site (Fig. 6) and have shown inhibition activity against PR (Forli et al. 2014). The pool of identified allosteric fragment binders was further expanded by crystallographic screening a library of 68 brominated fragments (Tiefenbrunn et al. 2014). Docking fragments can be a challenging task due to their small size and the

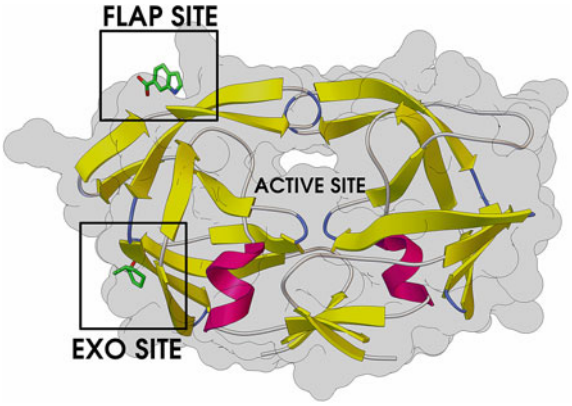



**Fig. 5** HIV-1 protease structure. Protein surface is shown as semitransparent MSMS (Sanner et al. 1996) surface and secondary structure. The location of the potential allosteric sites (Perryman et al. 2010b), flap and exo sites, is shown with boxes and the original fragments with which they were identified (1F1 and 4D9, respectively)

limited number of interactions that they can establish with the target structure. In fact, a smaller fragment in a given binding site can adopt more orientations than a drug-sized ligand, resulting in multiple poses to be evaluated and scored. This makes the identification of hits more sensitive to approximations in the scoring function, and it make the docking more prone to the generation of false positives. Also, as mentioned above, water molecules can have a much greater influence on the binding of fragments, as confirmed by our study on water-mediated docking (Forli and Olson 2012). From the perspective of chemical diversity, fragment libraries can be considered diversity libraries, where combinations of a relatively small number of molecules (1–10K compounds) cover a large chemical space (Chen and Shoichet 2009). Another common approach to build a diversity library is to remove structural redundancy, i.e., removing ligands with similar structure, and maximizing structural and chemical diversity of the set. A popular diversity set that has been used in many virtual screenings is the one provided by the Developmental Therapeutics Program (DTP) (NIH) of the National Cancer Institutes in USA [NCI diversity set (Holbeck 2004)] that was generated by using pharmacophoric models (NIH2). Early on, we used the NCI diversity set in combination with known PR inhibitors to evaluate docking enrichment factors on a screening of a small library of 1771 ligands against multiple mutants of PR. (Chang et al. 2007). An orthogonal approach to reduce the number of compounds in a library is to build focused libraries. Focused libraries are built using a scaffold or an entire ligand as a seed, then identifying compounds with high degree of structural similarity or presenting similar physicochemical properties (bioisosteres). Criteria involved in the generation of a focused library and the exploration of local chemical space around a seed can vary, depending on the main goal (e.g., increasing activity, creating pharmacological properties, or overcoming drug resistance). A small focused library

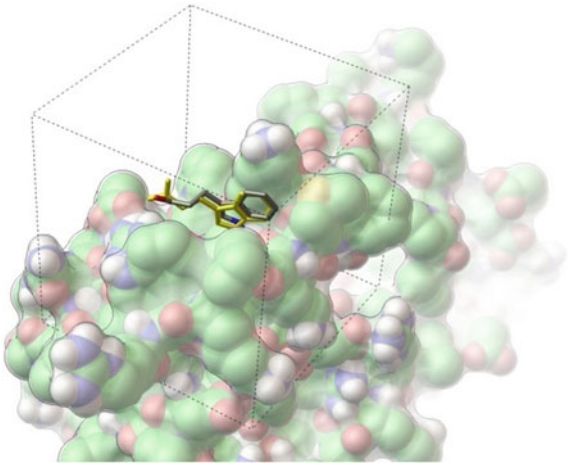


**Fig. 6** Fragment docking results on HIV-1 protease showing docked (*gray*) versus experimental (*yellow*) poses (PDB: *4ejl*) (Tiefenbrunn et al. 2013). The surface of the protein is shown with MSMS (Sanner et al. 1996) surface colored by atom types. The protein volume analyzed during dockings is shown as *dashed lines*

was built to screen for congeners of the 1F1 hit fragment previously identified (Perryman et al. 2010b). One of the fragments predicted by docking was confirmed by differential scanning fluorimetry (DSF) and showed a low micromolar binding constants to both apo-PR and pepstatin-bound PR, assayed by backscattering interferometry (BSI) (Baksh et al. 2011). Crystallographic results helped elucidate the dynamic mechanism of PR flaps and showed that docking predicted the experimental pose of the fragment with high accuracy (Tiefenbrunn et al. 2013) (Fig. 6).

Another way of building focused libraries is by virtual synthetic chemistry, where a validated synthetic pathway is simulated in silico. Once key steps of the synthesis are identified, a large number of commercially available reagents are combined to virtually enumerate all possible synthesizable derivatives. Such a library can be then screened computationally to prioritize the synthesis of derivatives by predicting their binding with the target, accelerating the drug design process.

HTVS experiments also provide a framework for exploring more in-depth structural aspects of the target, such as target flexibility or mutational variability. Flexibility is a crucial aspect (Heal et al. 2012) of proteins such as PR or IN, in which it plays an important role in their catalytic activity. While small conformational changes in residue side chains can be explored during a single docking, to date, dedicated calculations are required to simulate major conformational changes involving protein backbone (Fig. 7) or domain and sub-domain motions. Computational methods such as MD simulations and normal mode analysis (NMA) have been routinely applied to predict conformational changes. Protein conformations sampled using these methods can be used to perform parallel dockings in which the

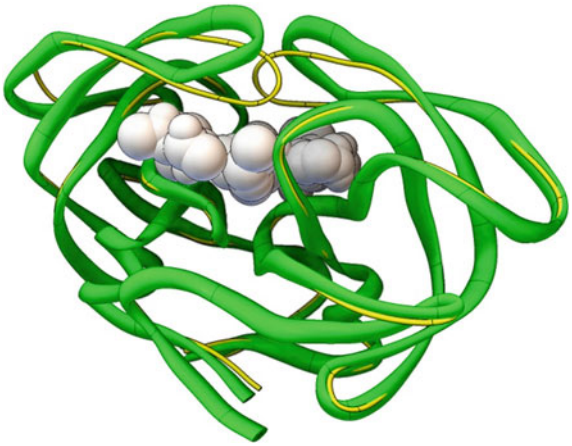


**Fig. 7** HIV-1 protease conformations showing unbound open conformation (*yellow ribbons,* PDB: *2pc0*) (Heaslet et al. 2007) overlapped to closed conformation (*green ribbons*, PDB: *3kf0*) (Perryman et al. 2010b) with the broad-based inhibitor TL-3 (*white spheres*) bound in the active site

ability of a ligand to bind is evaluated across a more realistic range of states of the target structures. Such a method has been termed the relaxed complex (RC) scheme (Lin et al. 2002). The study of the dynamic profile of a target also enables discovery and exploitation of intermediate or transient states that are difficult to identify in crystal structures. However, a significant side effect of sampling multiple target states is that it is possible to easily generate hundreds of target conformations, especially for unstructured loop regions. Despite the availability of methods to recapitulate the information contained in multiple conformations into a single state (Österberg et al. 2002; Cosconati et al. 2012), usually ligand libraries are docked separately against each target conformation. For this reason, having tools for reducing the multiplicity of structures, while keeping the same amount of conformational information, is essential. An example of such a tool successfully applied in the context of HIV structures is QR factorization (Amaro et al. 2008), a method able identify the most informative states in a given MD trajectory reducing structural redundancy while conserving a comparable dynamic profile description.

Structural data collected on wild-type PR and mutants cover a fairly large spectrum of conformational states in which the protein dimer can be found. MD simulations on PR studied the effect of mutations on the affinity of inhibitors used in the treatment of AIDS (Perryman et al. 2006). Results of MD analysis correlated flap movements with drug resistance. In particular, V82F/I84V mutations affect the equilibrium between semi-open and closed states with respect to the wild type, suggesting that dynamic perturbation effects are a likely cause for the mechanism of resistance (Perryman et al. 2004). Another intriguing finding of the study was the possibility of an allosteric site for inhibiting PR activity, resulting from the anti-correlation relationship between the fluctuations in the active site flaps and a distal

region, located in proximity to Gly40 and Gln61. This hypothesis was further explored with restrained dynamics simulations (Perryman et al. 2006) and finally validated with the crystallographic fragment screening that led to the identification of exposit binders (Perryman et al. 2010b). Eventually, site-specific mutation experiments confirmed that flap dynamics and corresponding elbow movements are crucial for the mechanism of action (Mittal et al. 2012).

MD simulations were also applied in a study on the flexibility of IN CCD. We used a restrained MD simulation protocol to better model the DDE motif geometry and the proper monodentate coordination of the catalytic magnesium ions (Perryman et al. 2010a). The study was performed combining molecular mechanics for modeling most of the enzyme structure and quantum chemistry calculations to describe the charge distribution on the catalytic residues and the $Mg^{2+}$ ions. The RC scheme was applied to extract multiple snapshots describing different protein conformations to be used in docking studies. The resulting models rationalized the effect of mutations related to the development of drug resistance on the binding of the IN strand transfer inhibitor (INSTI) raltegravir. In particular, the study compared wild-type and G140S/Q148H mutations, and their influence on the dynamic profile of the catalytic region of the enzyme.

Amino acid mutations exist in every HIV protein (Leslie et al. 2004), and the mutational explorations of the virus enable its rapid evolution of drug resistance. Computational methods can be used to analyze and correlate mutational data on patients under HIV drug therapy (Chang and Torbett 2011), and enable rationalization of the complex mechanisms of drug resistance, predicting their effects by modeling them in silico (Belew and Chang 2006). Most likely, a mutant protein will have a different interaction pattern with a given inhibitor and will present an altered dynamic profile, thereby increasing the dimensional complexity of any modeling efforts. For this reason, a tremendous amount of computing power is required to handle the escalating numbers of dockings of ligand virtual libraries against wild-type and mutation variants of different biological targets, a task whose complexity exceeds even large computational resources, such as *discrete* supercomputers. For this reason, in 2000, we started the FightAIDS@Home project (FA@H 2014), the first biomedical distributed computing project. The project was based on software from Entropia Inc, designed to use idle computing cycles from personal computers from donors worldwide. In 2005, FA@H joined the IBM World Community Grid (WCG) (Grid 2014), a public distributed computing infrastructure dedicated to computing projects that will benefit humanity. IBM Corporation has provided hardware, software, and technical services and expertise to build the WCG infrastructure.

Initially, FightAIDS@Home allowed us to test virtual screenings of a diverse ligand library consisting of compounds available from the NCI Diversity Set (NIH2 2014) and known protease inhibitors (Chang et al. 2007). The library was docked against 268 different wild-type and mutants of HIV-1 and HIV-2 protease structures, performing almost 500K different dockings and $10^{15}$ energy evaluations. Results from such a large-scale docking validated the docking protocol, showing that the scoring function was able to discriminate effectively between the known

binders and other ligands in the library. Furthermore, the screening enabled an efficient search for compounds active against a broad range of protease mutants and sub-types, all at the same time. Interaction analysis identified several structural determinants of the inhibition across the different structures. Performing the same calculation on computational resources available in our laboratory at the time would have been prohibitive.

The FA@H project spearheaded large-scale dockings, which benefited the whole docking community. In fact, tools and protocols developed from it have been shared and distributed all over the world with laboratories using AutoDock and related software. Among the several milestones achieved, in September 2009, FA@H received 100,000 CPU years of computation from WCG, while in June 2014, we approached 300,000 CPU years. In July of the same year, FA@H was the first WCG project to be run on android smartphones, exploiting the ever-increasing mobile device processing time. Today, IBM has over 2 million donated CPUs registered in its World Community Grid made available to researchers such as us.

Dealing with the throughput of results from FA@H and their consequent analysis, processing and archiving in itself poses a significant challenge, requiring the design of new tools. We are implementing database solutions to design and process FA@H experiments, managing results that span several years and billions of docking results. Initially, the FA@H project started with PR structures and has been extended to include IN and RT, while other target structures are under evaluation to be added. The successful application of AutoDock on FA@H has spawned several other WCG projects which use the docking codes developed in our laboratory to find new drugs for numerous therapeutically relevant targets, including malaria, dengue fever, neuroblastoma, and influenza (Grid 2014).

## 5 Multi-Scale and Mesoscale Modeling

Multi-scale biological models of HIV that span from atoms to the entire virion and its cellular interactions can greatly improve our understanding of the mechanisms of viral infection, proliferation, and drug resistance. However, experimental methods do not yet exist to directly observe, visualize, or model the mesoscale (the intermediate scale of $10^{-7}$–$10^{-8}$ m) bridging molecular and cellular biology. To address this need, we have developed a software framework called cellPACK (Johnson et al. 2014), which integrates multiple types of data across scales into comprehensive 3-D spatial models. We have begun to use this tool to build rigorous molecular models of HIV-1 from structural components and systems of biological data (Fig. 8). As structural, proteomic, genomic, and interatomic data on HIV continue to expand, we have the opportunity to synthesize an integrated structural picture of the complete virus and its life cycle.

The ability to automatically create structural models of the intact HIV virion using the latest data will enable researchers to generate input to a variety of simulation methods at different time and spatial scales. To this point, the generation of

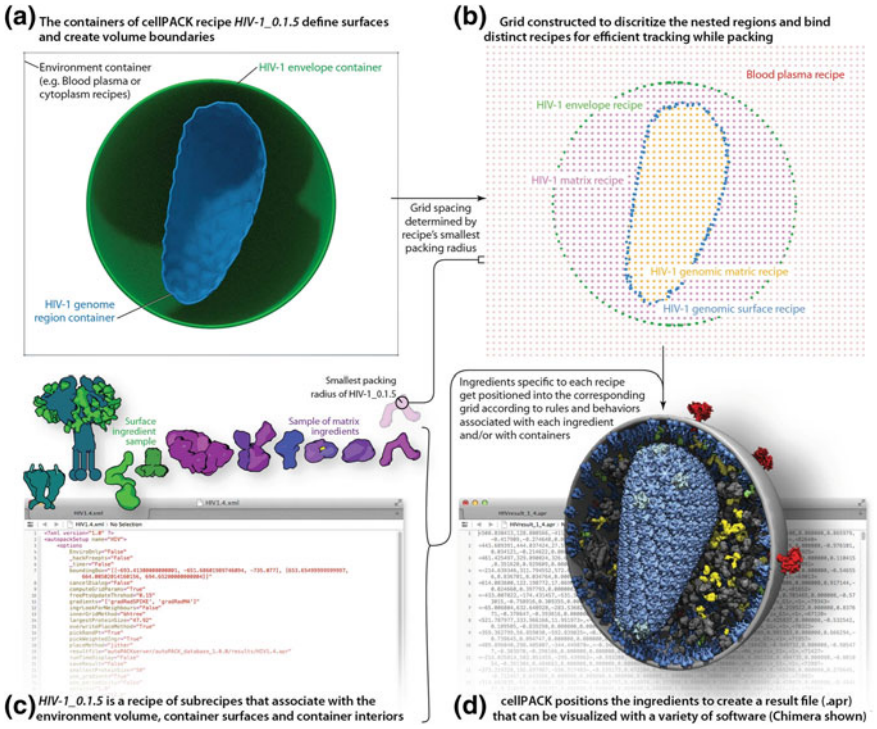



**Fig. 8** CellPACK uses a grid-based method to efficiently track ingredients as they are packed into recipe bounding containers. **a** Polyhedral mesh containers create boundaries and regions (e.g. HIV-1 envelope and matrix) restricting and compartmentalizing the packing of ingredients to interior and exterior regions of the capsid, for example. **b** A grid automatically spaced based on the smallest ingredient is calculated on the volume to track ingredients packing. **c** The full recipe of ingredients is packed (D) The results can be visualized and manipulated on a variety of visualization or 3-D animation software by means of a portable file format (adapted from (Johnson et al. 2014))

such models has been a serious bottleneck in simulation workflow. Importantly, the automated nature of the model generation in cellPACK enables the construction of large numbers of unique models, each consistent with the known range and distribution of molecular ingredients. This, in turn, enables the statistical analysis of a large number of simulations and the generation of multiple testable hypotheses.

## 6 Conclusions

Computational methods have provided fundamental support to HIV research since the initial structural studies, helping to unravel details of HIV biology. Computational models have proved to be a powerful tool to analyze and understand the

impact of mutations, and to overcome their structural and functional influence in drug resistance. With the availability of structural data, in silico experiments have been instrumental in exploiting and improving interactions between drugs and viral targets, such as HIV protease, reverse transcriptase, and integrase. Remarkable results have been achieved in understanding the biology of HIV as well as the capabilities and limitations of structure-based drug design modeling techniques. Yet, we still have a long way to go.

With the combination of increasing computational power and the large amount of structural and biological data continuing to accumulate, multi-scale approaches are becoming more feasible. Such models will allow large-scale modeling of large molecular complexes to whole virions. We therefore anticipate a shift from individual atomistic models toward mesoscale models where the structure and the interactions of large environments of molecular complexes will be modeled.

The identification and validation of new targets has significantly benefited from molecular docking and other computational methods. In the future, they will play a critical role in the study of protein interactions within these larger holistic models. Within this context, large multi-disciplinary efforts are a necessary approach. Indeed, in 2012, we formed the HIV interaction and Viral Evolution (HIVE) Center, whose goal is to characterize at the atomic level the structural and dynamic relationships between interacting macromolecules in the HIV life cycle in order to understand the mechanistic evolution of drug resistance. The center involves 13 individual researcher groups from seven institutions around the USA utilizing different methodologies in a single joint effort. This collaborative framework provides the expertise required to investigate the complexity of the HIV system with a multifaceted and versatile approach. From a computational perspective, the center provides the essential capabilities of in vitro and in vivo assays for rapid and accurate experimental validations of hypotheses generated in silico. Similarly, it is important to have agile computational tools that are able to generate a high-quality model in support of experimental results and that can be improved, in an iterative fashion, from those results.

In conclusion, computational methods will continue to be essential in the accumulation and interpretation of the tremendous amount of information that is gathered on the biological nature and human health impacts of HIV. Therefore, in the future, we expect to face new computational challenges and to have an even greater role in the design of more holistic approaches to the HIV treatment and ultimately to its eradication.

**Acknowledgments** We thank IBM World Community Grid for the computational resource support provided to the FightAIDS@Home project. This work was supported by NIH R01 GM073087 and P50 GM103368 to AJO.

## References

Adachi M, Ohhara T, Kurihara K, Tamada T, Honjo E, Okazaki N, Arai S, Shoyama Y, Kimura K, Matsumura H (2009) Structure of HIV-1 protease in complex with potent inhibitor KNI-272 determined by high-resolution X-ray and neutron crystallography. Proc Nat Acad Sci 106 (12):4641–4646

Ala PJ, DeLoskey RJ, Huston EE, Jadhav PK, Lam PY, Eyermann CJ, Hodge CN, Schadt MC, Lewandowski FA, Weber PC (1998) Molecular recognition of cyclic urea HIV-1 protease inhibitors. J Biol Chem 273(20):12325–12331

Alvizo O, Mittal S, Mayo SL, Schiffer CA (2012) Structural, kinetic, and thermodynamic studies of specificity designed HIV-1 protease. Protein Sci 21(7):1029–1041

Amaro RE, Baron R, McCammon JA (2008) An improved relaxed complex scheme for receptor flexibility in computer-aided drug design. J comput Aided Mol Des 22(9):693–705

Baksh MM, Kussrow AK, Mileni M, Finn M, Bornhop DJ (2011) Label-free quantification of membrane-ligand interactions using backscattering interferometry. Nat Biotechnol 29(4): 357–360

Baldwin ET, Bhat TN, Gulnik S, Liu B, Topol IA, Kiso Y, Mimoto T, Mitsuya H, Erickson JW (1995) Structure of HIV-1 protease with KNI-272, a tight-binding transition-state analog containing allophenylnorstatine. Structure 3(6):581–590

Belew RK, Chang MW (2006) Modeling recombination's role in the evolution of HIV drug resistance. In: Rocha LM, Bedau M, Floreano D, Goldstone R, Vespignani A, Yaeger L (eds) Artificial life X: the tenth international conference on the synthesis and simulation of living systems. MIT Press, Citeseer, pp 98–104

Binkowski TA, Naghibzadeh S, Liang J (2003) CASTp: computed atlas of surface topography of proteins. Nucleic Acids Res 31(13):3352–3355

Capra JA, Laskowski RA, Thornton JM, Singh M, Funkhouser TA (2009) Predicting protein ligand binding sites by combining evolutionary sequence conservation and 3D structure. PLoS Comput Biol 5(12):e1000585

Chang MW, Torbett BE (2011) Accessory mutations maintain stability in drug-resistant HIV-1 protease. J Mol Biol 410(4):756–760

Chang MW, Lindstrom W, Olson AJ, Belew RK (2007) Analysis of HIV wild-type and mutant structures via in silico docking against diverse ligand libraries. J Chem Inf Model 47(3): 1258–1262

Chen Y, Shoichet BK (2009) Molecular docking and ligand specificity in fragment-based inhibitor discovery. Nat Chem Biol 5(5):358–364

Christ F, Voet A, Marchand A, Nicolet S, Desimmie BA, Marchand D, Bardiot D, Van der Veken NJ, Van Remoortel B, Strelkov SV (2010) Rational design of small-molecule inhibitors of the LEDGF/p75-integrase interaction and HIV replication. Nat Chem Biol 6(6):442–448

Cosconati S, Forli S, Perryman AL, Harris R, Goodsell DS, Olson AJ (2010) Virtual screening with AutoDock: theory and practice. Expert Opin Drug Discov 5(6):597–607

Cosconati S, Marinelli L, Di Leva FS, La Pietra V, De Simone A, Mancini F, Andrisano V, Novellino E, Goodsell DS, Olson AJ (2012) Protein flexibility in virtual screening: the BACE-1 case study. J Chem Inf Model 52(10):2697–2704

Engelman A, Cherepanov P (2012) The structural biology of HIV-1: mechanistic and therapeutic insights. Nat Rev Microbiol 10(4):279–290

FightAids@Home FA@H project page (2014) http://fightaidsathome.scripps.edu/. Accessed 21 July 2014

Forli S, Olson AJ (2012) A force field with discrete displaceable waters and desolvation entropy for hydrated ligand docking. J Med Chem 55(2):623–638

Forli S, Tiefenbrunn T, Stout CD, Olson AJ (2014) Allosteric inhibitors of HIV protease, unpublished results, TSRI

Gallicchio E, Lapelosa M, Levy RM (2010) Binding energy distribution analysis method (BEDAM) for estimation of Protein-Ligand binding affinities. J Chem Theor Comput 6 (9):2961–2977

Gallicchio E, Deng N, He P, Wickstrom L, Perryman AL, Santiago DN, Forli S, Olson AJ, Levy RM (2014) Virtual screening of integrase inhibitors by large scale binding free energy calculations: the SAMPL4 challenge. J Comput Aided Mol Des 28(4):475–490

Goodsell DS, Olson AJ (1990) Automated docking of substrates to proteins by simulated annealing. Proteins Struct Funct Bioinform 8(3):195–202

Grid IWC WCG (2014) http://www.worldcommunitygrid.org/. Accessed 21 July 2014

Grzesiek S, Bax A, Nicholson LK, Yamazaki T, Wingfield P, Stahl SJ, Eyermann CJ, Torchia DA, Hodge CN (1994) NMR evidence for the displacement of a conserved interior water molecule in HIV protease by a non-peptide cyclic urea-based inhibitor. J Am Chem Soc 116(4): 1581–1582

Hajduk PJ, Greer J (2007) A decade of fragment-based drug design: strategic advances and lessons learned. Nat Rev Drug Discov 6(3):211–219

Hall HI, Song R, Rhodes P, Prejean J, An Q, Lee LM, Karon J, Brookmeyer R, Kaplan EH, McKenna MT (2008) Estimation of HIV incidence in the United States. JAMA 300(5): 520–529

Halperin I, Ma B, Wolfson H, Nussinov R (2002) Principles of docking: an overview of search algorithms and a guide to scoring functions. Proteins Struct Funct Bioinform 47(4):409–443

Hare S, Vos AM, Clayton RF, Thuring JW, Cummings MD, Cherepanov P (2010) Molecular mechanisms of retroviral integrase inhibition and the evolution of viral resistance. Proc Nat Acad Sci 107(46):20057–20062

Harris R, Olson AJ, Goodsell DS (2008) Automated prediction of ligand-binding sites in proteins. Proteins Struct Funct Bioinform 70(4):1506–1517

Heal J, Jimenez-Roldan JE, Wells SA, Freedman R, Römer RA (2012) Inhibition of HIV-1 protease: the rigidity perspective. Bioinformatics 28(3):350–357

Heaslet H, Rosenfeld R, Giffin M, Lin Y-C, Tam K, Torbett BE, Elder JH, McRee DE, Stout CD (2007) Conformational flexibility in the flap domains of ligand-free HIV protease. Acta Crystallogr Sect. D: Biol Crystallogr 63(8):866–875

Hodge CN, Aldrich PE, Bacheler LT, Chang C-H, Eyermann CJ, Garber S, Grubb M, Jackson DA, Jadhav PK, Korant B (1996) Improved cyclic urea inhibitors of the HIV-1 protease: synthesis, potency, resistance profile, human pharmacokinetics and X-ray crystal structure of DMP 450. Chem Biol 3(4):301–314

Holbeck S (2004) Update on NCI in vitro drug screen utilities. Eur J Cancer 40(6):785–793

Jadhav PK, Ala P, Woerner FJ, Chang C-H, Garber SS, Anton ED, Bacheler LT (1997) Cyclic urea amides: HIV-1 protease inhibitors with low nanomolar potency against both wild type and protease inhibitor resistant mutants of HIV. J Med Chem 40(2):181–191

Johnson G, Goodsell DS, Autin L, Forli S, Sanner M, Olson AJ (2014) FD 169: 3D molecular models of whole HIV-1 virions generated with cellPACK. Faraday Discuss 169:23–44

King NM, Prabu-Jeyabalan M, Bandaranayake RM, Nalam MN, Nalivaika EA, Ael Özen, Tr Haliloğlu, NeK Yılmaz, Schiffer CA (2012) Extreme Entropy-Enthalpy Compensation in a Drug-Resistant Variant of HIV-1 Protease. ACS Chem Biol 7(9):1536–1546

Kiso Y, Matsumoto H, Mizumoto S, Kimura T, Fujiwara Y, Akaji K (1999) Small dipeptide-based HIV protease inhibitors containing the hydroxymethylcarbonyl isostere as an ideal transition-state mimic. Pept Sci 51(1):59–68

Kuntz ID, Blaney JM, Oatley SJ, Langridge R, Ferrin TE (1982) A geometric approach to macromolecule-ligand interactions. J Mol Biol 161(2):269–288

Lange-Savage G, Berchtold H, Liesum A, Budt KH, Peyman A, Knolle J, Sedlacek J, Fabry M, Hilgenfeld R (1997) Structure of HOE/BAY 793 complexed to Human Immunodeficiency Virus (HIV-1) protease in two different crystal forms structure/function relationship and influence of crystal packing. Eur J Biochem 248(2):313–322

Laurie AT, Jackson RM (2005) Q-SiteFinder: an energy-based method for the prediction of protein–ligand binding sites. Bioinformatics 21(9):1908–1916

Leslie A, Pfafferott K, Chetty P, Draenert R, Addo M, Feeney M, Tang Y, Holmes E, Allen T, Prado J (2004) HIV evolution: CTL escape mutation and reversion after transmission. Nat Med 10(3):282–289

Liang S, Zhang C, Liu S, Zhou Y (2006) Protein binding site prediction using an empirical scoring function. Nucleic Acids Res 34(13):3698–3707

Lin J-H, Perryman AL, Schames JR, McCammon JA (2002) Computational drug design accommodating receptor flexibility: the relaxed complex scheme. J Am Chem Soc 124 (20):5632–5633

Lu C, Li AP (2010) Enzyme inhibition in drug discovery and development: the good and the bad. Wiley, New York

Marshall GR (1987) Computer-aided drug design. Ann Rev Pharmacol Toxicol 27(1):193–213

Mittal S, Cai Y, Nalam MN, Bolon DN, Schiffer CA (2012) Hydrophobic core flexibility modulates enzyme activity in HIV-1 protease. J Am Chem Soc 134(9):4163–4168

Mobley DL, Liu S, Lim NM, Wymer KL, Perryman AL, Forli S, Deng N, Su J, Branson K, Olson AJ (2014) Blind prediction of HIV integrase binding from the SAMPL4 challenge. J Comput Aided Mol Des 28(4):327–345

Moon JB, Howe WJ (1991) Computer design of bioactive molecules: a method for receptor-based de novo ligand design. Proteins Struct Funct Bioinform 11(4):314–328

Morris GM, Goodsell DS, Halliday RS, Huey R, Hart WE, Belew RK, Olson AJ (1998) Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. J Comput Chem 19(14):1639–1662

Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, Olson AJ (2009) AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. J Comput Chem 30(16):2785–2791

Navia MA, Fitzgerald PM, McKeever BM, Leu CT, Heimbach JC, Herber WK, Sigal IS, Darke PL, Springer JP (1989) Three-dimensional structure of aspartyl protease from human immunodeficiency virus HIV-1. Nature 337(6208):615–620. doi:10.1038/337615a0

Nicholson LK, Yamazaki T, Torchia DA, Grzesiek S, Bax A, Stahl SJ, Kaufman JD, Wingfield PT, Lam PY, Jadhav PK (1995) Flexibility and function in HIV-1 protease. Nat Struct Biol 2 (4):274–280

NIH Developmental Therapeutics Program (2014) http://dtp.nci.nih.gov/. Accessed 11 June 2014

NIH2 DTP NCI Diversity set IV (2014) http://dtp.nci.nih.gov/branches/dscb/div2_explanation.html. Accessed 10 June 2014

Österberg F, Morris GM, Sanner MF, Olson AJ, Goodsell DS (2002) Automated docking to multiple target structures: incorporation of protein mobility and structural water heterogeneity in AutoDock. Proteins Struct Funct Bioinform 46 (1):34–40

Perryman AL, Lin JH, McCammon JA (2004) HIV-1 protease molecular dynamics of a wild-type and of the V82F/I84V mutant: possible contributions to drug resistance and a potential new target site for drugs. Protein Sci 13(4):1108–1123

Perryman AL, Lin JH, McCammon JA (2006) Restrained molecular dynamics simulations of HIV-1 protease: the first step in validating a new target for drug design. Biopolymers 82(3):272–284

Perryman AL, Forli S, Morris GM, Burt C, Cheng Y, Palmer MJ, Whitby K, McCammon JA, Phillips C, Olson AJ (2010a) A dynamic model of HIV integrase inhibition and drug resistance. J Mol Biol 397(2):600–615

Perryman AL, Zhang Q, Soutter HH, Rosenfeld R, McRee DE, Olson AJ, Elder JE, David Stout C (2010b) Fragment-Based Screen against HIV Protease. Chem Biol Drug Des 75(3):257–268

Perryman AL, Santiago DN, Forli S, Santos-Martins D, Olson AJ (2014) Virtual screening with AutoDock Vina and the common pharmacophore engine of a low diversity library of fragments and hits against the three allosteric sites of HIV integrase: participation in the SAMPL4 protein–ligand binding challenge. J Comput Aided Mol Des 28(4):429–441

Sanner MF, Olson AJ, Spehner JC (1996) Reduced surface: an efficient way to compute molecular surfaces. Biopolymers 38(3):305–320

Suresh CH, Vargheese AM, Vijayalakshmi KP, Mohan N, Koga N (2008) Role of structural water molecule in HIV protease-inhibitor complexes: a QM/MM study. J Comput Chem 29 (11):1840–1849

Tiefenbrunn T, Forli S, Baksh MM, Chang MW, Happer M, Lin Y-C, Perryman AL, Rhee J-K, Torbett BE, Olson AJ (2013) Small molecule regulation of protein conformation by binding in the flap of HIV protease. ACS Chem Biol 8(6):1223–1231

Tiefenbrunn T, Forli S, Happer M, Gonzalez A, Tsai Y, Soltis M, Elder JH, Olson AJ, Stout CD (2014) Crystallographic fragment-based drug discovery: use of a brominated fragment library targeting HIV protease. Chem Biol Drug Des 83(2):141–148

Trott O, Olson AJ (2010) AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem 31(2): 455–461

Volkamer A, Kuhn D, Rippmann F, Rarey M (2012) DoGSiteScorer: a web server for automatic binding site prediction, analysis and druggability assessment. Bioinformatics 28(15): 2074–2075

Wang Y-X, Freedberg DI, Wingfield PT, Stahl SJ, Kaufman JD, Kiso Y, Bhat TN, Erickson JW, Torchia DA (1996) Bound water molecules at the interface between the HIV-1 protease and a potent inhibitor, KNI-272, determined by NMR. J Am Chem Soc 118(49):12287–12290

Wang R, Lu Y, Wang S (2003) Comparative evaluation of 11 scoring functions for molecular docking. J Med Chem 46(12):2287–2303

Warren GL, Andrews CW, Capelli A-M, Clarke B, LaLonde J, Lambert MH, Lindvall M, Nevins N, Semus SF, Senger S (2006) A critical assessment of docking programs and scoring functions. J Med Chem 49(20):5912–5931

Weber IT (1990) Evaluation of homology modeling of HIV protease. Proteins Struct Funct Bioinform 7(2):172–184

Whiting M, Muldoon J, Lin Y-C, Silverman SM, Lindstrom W, Olson AJ, Kolb HC, Finn M, Sharpless KB, Elder JH (2006) Inhibitors of HIV-1 protease by using in situ click chemistry. Angew Chem Int Ed 45(9):1435–1439

http://www.springer.com/978-3-319-18517-0

# The Future of HIV-1 Therapeutics
## Resistance Is Futile?

Torbett, B.; Goodsell, D.S.; Richman, D. (Eds.)

2015, X, 254 p., Hardcover
ISBN: 978-3-319-18517-0