Area of Science Designation: 10 Quantitative and Computational Biology

NOVEL PARADIGMS FOR DRUG DISCOVERY: COMPUTATIONAL MULTITARGET SCREENING

PROJECT DESCRIPTION

We will develop a novel computational approach to drug screening and identification that takes advantage of state of the art protein folding and small molecule docking with dynamics methods. Furthermore, we will benchmark and improve our method via rigorous experimental assessment of prospective predictions.

APPROACH

The discovery of a new drug targeting a specific disease condition is usually initiated by finding ”hits” against ”target” proteins of interest using experimental high throughput screening (HTS) against a large chemical compound library. The *in vitro* hits are then assessed in an *in vivo* system whereupon active compounds are considered ”leads” to a drug. Finally the long and arduous process of three phases of clinical trials is undertaken to obtain approval from a governing agency (FDA in the US).

This traditional process is inefficient since the blinded screens result in a vast number of hits and leads failing to be effective and safe in humans. The major pharmaceutical companies (Pharma) therefore estimate that the cost of developing a new drug to market is in the order of $1.5 billion dollars including the cost of failures. Pharma sets prices to derive profit beyond this tremendous overhead, and as a result therapeutics are often out of reach to those who need it most. In addition, for many infectious diseases of the third world, there is little or no profit too be made so many potential drug targets these diseases are largely ignored in industry.

Over the past several years, we have developed an infrastructure to not only predict protein structure but to successfully dock small molecules to actual and predicted protein structures, using similar principles for energetics and exploring the search space. In this application, we seek to develop our methods into a novel pipeline for rapid drug discovery that can be applied to any protein target.

Our approach is to computationally curate an intelligently compiled compound library, select for inhibitory function in a set of target disease proteins, and avoid side effect activity in an alternate set of host proteins. We posit that by using an unblinded approach to computational drug screening, we can greatly catalyze the efficiency and success of drug discovery, thereby making pharmaceutical development viable across many private and public research settings, making the pharmacological paradigm of medical treatment viable across all socioeconomic statuses, and simultaneously enabling personalised medicine.

Computational methods

We have developed a new computational paradigm for the discovery of potential lead inhibitors that is based on a combination of several tenets: (*i*) Incorporate protein side chains and main chain dynamics during the docking process to more accurately evaluate binding affinities. (*ii*) Effectively sample the conformational space for compounds with torsional flexibility by separately docking the rigid fragments, translated and rotated along a finite lattice, with maximum parsimony annealing. (*iii*) Apply knowledge based functions to discriminate the optimal conformation for a compound target pair, and then compare across compounds to select the optimal compound for the target. (*iv*) Select single inhibitors that bind to multiple protein disease related targets simultaneously. (*v*) Filter compounds which bind to important host and commensal microbial proteins. (*vi*) Use a screening library consisting of drug and drug-like compounds (i.e., those that have already passed some form of preliminary human evaluation).

Tenets 1 and 2 (docking with dynamics) differentiate our work from a previous large body of mostly unsuccessful computational approaches to drug discovery as in most previous work both the protein and the small molecules were considered to be rigid structures. The correlation between computed and experimentally measured binding energies is 0.35 when the protein and inhibitor are considered as rigid bodies and 0.88 when both the protein and inhibitor are allowed to be flexible (Jenwitheesuk & Samudrala, 2003[1]).

Each tenet increases the probability that a predicted compound will successfully inhibit the disease. Furthermore, screening with drug-like compounds specifically increases pharmacological viability. This new paradigm produces hits that will more expediently and predictably become lead compounds that can po-

[1] References are available via the biographical sketch.

tentially be developed further into viable drugs for all diseases. The tenets will each be represented by separable software and databases within an integrated suite that will be made freely and publicly available to accelerate drug discovery in public and private laboratories.

Comparison to traditional methods

We anticipate that bench verification in physical systems simulating the host, such as cell culture and model organisms, will never completely be replaced computational simulation. However, it can serve as a powerful filter to reduce the number of screening possibilities or to identify potential new targets for drug repositioning. Computational simulation allows simultaneous consideration of more factors than any one experimental simulation can assess. We can model the absorption, distribution, metabolism, and excretion systems (ADME). We can model nontarget binding in both human and commensal microbes, which induce untoward side effects. We can model positive side effects such as inhibitory binding of other disease targets which do commonly occur (as described by our recent review (Jenwitheesuk et al., 2008), work describing HIV-1 targets of minocycline (Jenwitheesuk & Samudrala, 2007), and work describing efficacy of HIV-1 drugs against CMV (Jenwitheesuk & Samudrala, 2005a)). We extensively use knowledge based approaches to inform the modelling along with 3D simulations.

Preliminary studies and successes

We have a history of success in developing predictive methods to aid therapeutic discovery. First, we created a freely available web server to select the optimal drug for HIV-1 patients based on the HIV-1 protease and reverse transcriptase sequence mutations, using both sequence and structural simulation methods assessing capacity for resistance versus susceptibility (Jenwitheesuk & Samudrala, 2003; Wang et al., 2004; Jenwitheesuk et al ., 2004; Jenwitheesuk et al., 2005; Jenwitheesuk & Samudrala, 2005b). Our web server has handled over 1,000 separate queries (<protinfo.compbio.washington.edu/pirspred/>). We also identified positive side effects explaining that the clinical observation of the AIDS related opportunistic pathogen CMV being cleared in certain HIV-1 targeting regimens is due to specific activity of amprenavir and indinavir against the CMV protease (Jenwitheesuk & Samudrala, 2005a), and that the mode of HIV-1 inhibition by the common antibiotic minocycline is through HIV-1 integrase (Jenwitheesuk & Samudrala, 2007).

Directly, a simpler earlier version of this drug discovery protocol was used in a prospective manner to computationally predict sixteen inhibitors for *Plasmodium falciparum*, of which six showed lethality at sub micromolar concentrations; further, we demonstrated the ability to predict inactivity, wherein all five negative prediction compounds displayed immeasurable effects on *P. falciparum* growth (Jenwitheesuk & Samudrala, 2005c; Jenwitheesuk et al., 2008). As well, we generalised this protocol and our protein structure prediction pipeline in a reverse engineering approach to design peptides which were experimentally demonstrated to prevent infectivity of dengue virus at the micromolar level (Costin, et al., *submitted*). We continued this approach to prospectively predict drugs to prevent infectious disease organisms by targeting the entire family of herpesviruses; we tested a single compound against representatives of the three subfamilies: herpes simplex virus-1 (HSV-1; alpha), cytomegalovirus (CMV; beta), and Kaposi's sarcoma associated herpesvirus (KSHV; gamma). This compound showed 10^{-5}M inhibition of viral growth (Figure 2) and similar dissociation constants to the respective proteases (Figure 3). Thus with a protocol drastically more simple and limited than that which we are proposing, we have already achieved orders of magnitude higher success at profoundly lower cost and time than Pharma HTS (Figure 1). Moreover, we demonstrate our commitment to making all data freely available for the advancement of medicine and science, by specifically displaying the identity of all hit compounds in all related publications.

Tenet rationale and methodologies developed by us.

(*i*) *Docking with protein target dynamics*. Biologically active proteins are in continuous motion, yet the majority of protein structure information is limited to the most stable form of a protein when crystallised in artificial conditions. Induced fit is a widely recognised challenge in computational drug screening, wherein the protein undergoes significant conformational changes upon ligand binding. As a consequence, traditional rigid protein-ligand docking is insufficient, and oftentimes misleading, for structure guided drug screening. Dynamics simulations increase the possibility of surveying a physiologically relevant conformation beyond using the static crystal structure alone. We identify a rigid fragment of the ligand to initiate docking to a binding region of the protein with high structural stability. The problem of induced fit is ad-

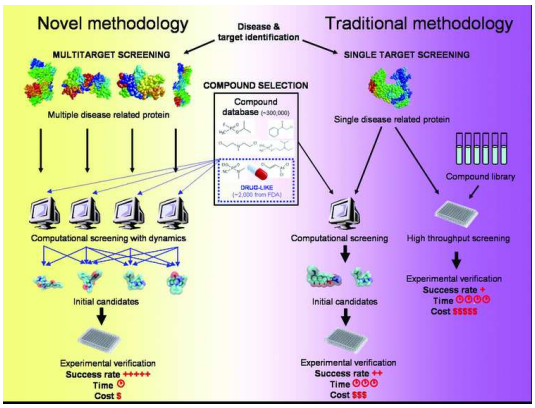

Figure 1: Our novel multitarget drug discovery protocol (left hand side) compared to traditional high throughput methods used by the pharmaceutical industry (right hand sides). We can effectively use our integrated screening pipeline to greatly increase the success rate, and reduce the cost and time taken for a clinically successful drug.

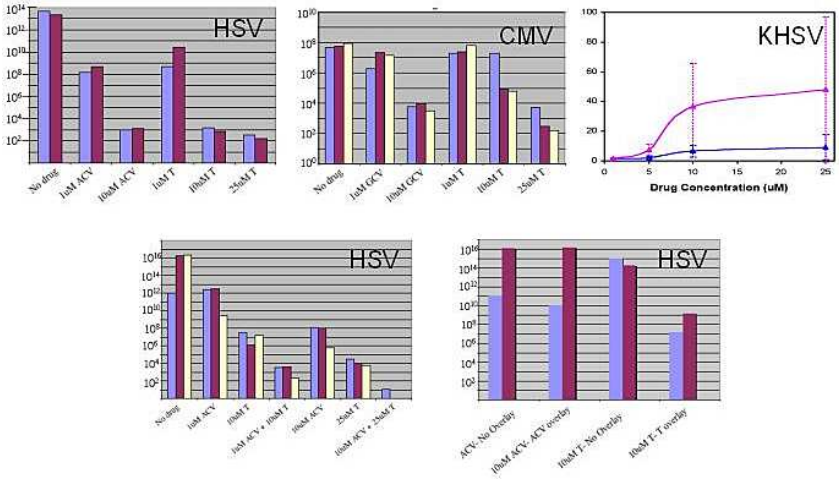

Figure 2: Inhibition of TMPyP4, our top predicted inhibitor, against HSV-1, CMV, and KSHV (top three panels). Cells were cultured in the absence or presence of TMPyP4, and antiherpes drugs acylovir or gancyclovir. Virus from the infected cells were then titered. The titers of each infection from 2 or 3 separate experiments are shown. Vero cells were infected with HSV-1 strain F. HFF cells were infected with CMV. Our inhibitor works as well or better than existing antiherpes drugs against HSV and CMV (top left two panels). Using a different assay, TMPyP4 also inhibited KSHV, a gamma herpesvirus (top right most panel). Note that all panels except the top right most panel display the amount of virus on a log scale whereas the KSHV data shows the fold reduction. Our computationally predicted broad spectrum human herpesvirus protease inhibitors is effective *in vitro* against members from all three classes and is comparable or better than antiherpes drugs. Our protease inhibitor acts synergistically with acylovir (a nucleoside analogue that inhibits replication) so much that it almost completely eliminates all virus when used together (bottom left panel). After passage for several cycles with both acylovir and TMPyP4, complete resistance to acylovir occurs, whereas our inhibitor still continues to be effective (bottom right panel).

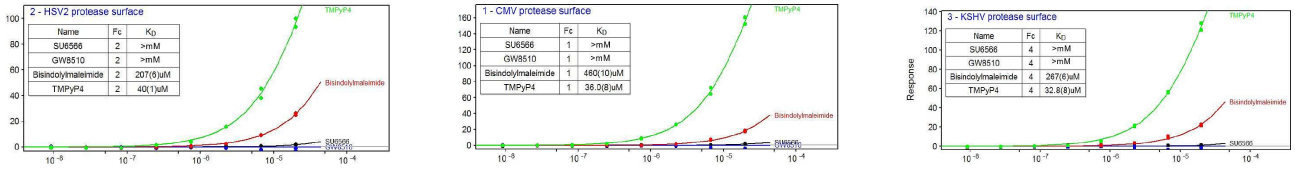

Figure 3: Disassociation constants (K_d) of our top ranking inhibitors for HSV, CMV, and KSHV experimentally determined using Surface Plasmon Resonance (SPR). The experiments were performed by Dr. David Myszka at the University of Utah. Four inhibitors were evaluated; our top ranking one (TMPyP4) clearly binds to all three proteases as predicted. Our second ranking inhibitor (Bisindolylmaleimide) binds with lower affinity than our top one, but also targets all three protease molecules. A third inhibitor, GW8510, predicted to bind fails (i.e., a misprediction). A fourth inhibitor, SU6566, predicted and verified to not bind was used as a negative control. The K_d of TMPyP4 is in the micromolar range which is consistent with our prediction of the inhibitor binding to herpesvirus protease monomers. The dimerisation constant of herpesvirus proteases is also in the micromolar range. Our prediction of the K_d of TMPyP4 binding to the protease dimer is in the nanomolar range, but since we predict it disrupt dimerisation, this nanomolar binding is observed only transiently. Our inhibitor prediction is completely consistent with the observed experimental data which is verified further by cell culture studies in Figure 2 and compared to existing antiherpes drugs.

dressed as the fragmented ligand is linked together to reconstruct the complete ligand and the protein is allowed to relax with torsional dynamics around this binding mode.

(*ii*) *Fragmented compound lattice docking*. Compounds are computationally fragmented at their rotatable bonds and then positioned at uniform grid points and rotations at the protein surface. Molecular surface grid points are drawn for each fragment and protein atom to produce an initial round of coarse grained sampling. Fragment poses are then filtered by discriminatory function score, then spatially clustered high scoring poses to identify regions forming favorable interactions. The optimisation of binding mode clusters is followed by fragment linking and complete ligand reconstruction. As the reconstruction proceeds, ligand restrained minimisation (energy dynamics) is performed to simulate induced fit of the flexible protein side chains and main chains upon binding.

(*iii*) *Knowledge based discriminatory functions*. We have recently developed a generalised discriminatory function score to select optimal poses for any type of ligand, within a margin of error which will always be sampled by the course lattice method. The function outperforms more than 20 other published ones in several docking decoy tests (Bernard & Samudrala, 2009). This function complements the residue specific all atom probability discriminatory function which has been the most widely successful function to select native like predicted protein structures for over ten years (Samudrala & Moult, 1998). Along with the above two tenets, this comprises my group's major contribution to the computational docking field.

(*iv*) *Multitargeting*. The most effective drugs in humans (e.g. aspirin or Gleevec) inevitably interact with and bind to multiple proteins, a feature that traditional models based on single target drugs fail to take into account. The multitarget approach is a necessary one because every drug has to be effective at its site of action (for example, HIV-1 protease inhibitors have to bind and inhibit the protease molecule) and has to be readily metabolised by the body (for example the cytochrome P450 (CYP450) enzymes, which are responsible for metabolising the majority of drugs). Computational screening for multitarget binding and inhibition is effective because it exploits the evolutionary fact that protein structure is conserved much more in nature than is function or sequence. Most importantly, inhibitor resistance is largely overcome by the exponentially decreased probability of resistant mutations simultaneously arising in genes encoding proteins corresponding to all targets (Figure 1).

(*v*) *Avoid side effects and maximise physiologic compatibility*. We minimise the chance of side effects by applying our docking protocol to the set of all host proteins and identifying those compounds that do not bind to essential proteins ("antitargets"). Additionally we will seek antitarget protein sequences for which structures are not known, using a functional profile-profile search strategy and proteins structure prediction. This specific design to avoid antitargets is physically equivalent to optimising dissociation for these proteins, and thereby represents an effective novel approach to avoid side effects.

(*vi*) *Use a screening library consisting of drug and drug like compounds*. Living organisms have evolved in comparable chemical environments containing similar sets of organic molecules. This shared evolutionary chemical context sets the stage for various organisms to use the same compounds to control different processes, making one molecule relevant to diverse physiological activity. The observation that structural folds are largely conserved, even when sequence and function are not, provides logical evidence that one compound can be an excellent initial candidate for many different protein targets. More directly, if we have a given compound that has gone through the FDA approval process for one indication, we may be able to reposition it for other indications more readily. Our own herpes and malarial inhibitor predictions are examples of such multitargeting compounds.

EVIDENCE OF INNOVATION

My undergraduate majors (from 1990-1993) were in the areas of computing science and genetics. Obtaining a double major in the early 1990s in these two seemingly disparate fields is indicative of my foresight and intellectual vision. I have tremendous experience in both basic theoretical and applied research (over 80 publications total) including a demonstrated ability and interest to work extensively with experimental groups, and mentor graduate and undergraduate students. My graduate and postdoctoral work was in the area of protein structure. Since joining the University of Washington (UW), I have continued to develop methods for improved, rapid, and automated prediction and design of protein and proteome structure, function, and interactions (described in detail in the "most significant achievement" section), nanobiotechnology, and drug discovery.

There are a few projects that particularly showcase my ability to “think outside the box” and go beyond my traditional background of computational structural biology and bioinformatics: Our inhibitor discovery protocol, which has been applied to a number of diseases, predicted a broad spectrum herpesvirus inhibitor that has been shown to work as predicted in cell culture against all three classes of herpesviruses, comparable or better than existing antiherpes drugs. This procedure has also been applied to discover multitarget inhibitors against malaria and dengue that have also been experimentally verified. In collaboration with researchers from Thailand, Japan, and India, we are developing a series of computational technologies that will enable us to reengineer pathways in rice to provide a full range of bioavailable nutrients. Further, we have computationally designed peptides capable of binding specific inorganic substrates (such as quartz and hydroxyapatite) and, again, experiments performed by collaborators in the UW Materials Science department indicate that our designed peptides behave exactly as predicted. Finally, we represent the computational core of a large national consortium to rebuild the tooth.

These projects not only provide evidence of my interest in a diverse range of biomedical problems that are particularly overdue for fresh perspectives, but also showcase my willingness to develop, extend, and apply novel computational methods to solve disparate problems and generate hypotheses that can be verified experimentally. For all projects that involve experimental verification, I am the primary investigator and the person responsible for the experimental design: I am the PI on awards to determine the binding affinities of our predicted inhibitor against proteases from three herpesviruses; to determine the toxicity and efficacy of our predicted inhibitor in mouse models of herpes; and to determine the binding affinity and specificities of computationally designed peptides capable of binding to inorganic substrates.

My work as a PI initially resulted in recognition from the Kinship Foundation, which awarded me the prestigious Searle Scholars Fellowship. My general research is now funded by multiple awards from the National Science Foundation and the National Institutes of Health. I was named one of the world’s top young innovators by MIT’s Technology Review magazine in 2003, and was selected in 2004 to present the UW New Investigator Science in Medicine Lecture which honors outstanding young faculty. I also received an NSF CAREER award in 2005. Both the TR100 and CAREER awards honour visionary thinking and potential for scientific leadership at present and in the future. In 2006, I was an NIH Director’s Pioneer Award Finalist (25/465 applicants selected as finalists) for an entirely different application. In 2008, I won the Alberta Heritage Foundation for Medical Research Visiting Scientist Award and was awarded honorary diplomas from the cities of Carsma and Yaoton, Peru, for my humanitarian work on developing new structure based methods for vaccine discovery.

DIFFERENCES FROM PAST WORK

My graduate and postdoctoral research was purely computational. The major difference between what I propose above, and what we currently do, is the development of an hybrid computational/bench pipeline encapsulating the entire process of drug design and discovery (i.e., not just inhibitor discovery). In that light alone, this project represents a marked difference from our previous work. Additionally, the molecular biology bench experimental techniques we use (described above) are different from any of the ones we normally deal with in structural biology (i.e., x-ray diffraction or NMR spectroscopy). The unifying theme however is the use of hybrid technologies parameterised on existing noisy or limited experimental data and computational simulations, and used in an iterative manner to “bootstrap” an initial set of probabilities and predictions to filter experimental data and generate more accurate predictions.

SUITABILITY FOR THE PIONEER AWARD PROGRAM

To address criticism about a limited amount of bench experience, I have built teams of experts to get all parts of the necessary laboratory work done: For example, we have synthesised, cloned, and expressed the proteases from three herpesviruses (HSV-1, CMV, and KSHV) as part of the determination of binding affinities of a novel broad spectrum herpesvirus inhibitor computationally predicted by us and shown to work better than existing antiherpes drugs in cell culture (see section on evidence of innovation above). Our success in the field of protein structure, function, and interaction prediction suggests that our approach is feasible. To construct the pipeline described above, our goal has been to perform proof-of-concept experiments using startup funding, and work our way to larger awards through the NIH. This process will take years. A NIH Director’s Pioneer award would give us the freedom to construct a high throughput drug discovery pipeline in a more rapid and directed manner.