DEPARTMENT OF BIOINFORMATICS
ALAGAPPA UNIVERSITY
(Reaccredited with 'A' Grade by NAAC)
KARAIKUDI-630 004, TAMIL NADU, INDIA

5th National Symposium cum Workshop on "Recent Trends in Structural Bioinformatics and Computer Aided Drug Design" (SBCADD'2013)

Workshop Course Materials

SCHRÖDINGER.

Introduction to Computer Aided Drug Design

Drug design, also sometimes referred to as rational drug design, is the inventive process of finding new medications based on the knowledge of the biological target. The drug is most commonly an organic small molecule which activates or inhibits the function of a biomolecule such as a protein which in turn results in a therapeutic benefit to the patient. In the most basic sense, drug design involves design of small molecules that are complementary in shape and charge to the bimolecular target to which they interact and therefore will bind to it. Drug design frequently but not necessarily relies on computer modeling techniques. This type of modeling is often referred to as computer-aided drug design.

In computational drug discovery we are using computational tools and softwares to simulate drug receptor interactions. In experimental based approach drugs were discovered through trial and error method. And also it was taking high R&D cost and also time consuming. Using computational drug discovery helps scientists to understand large insight to the drug receptor interactions. And also it helps to reduce the time and cost. Scientists are able to predict whether the molecule will be success or fail in the market.

There are two major types of drug design. The first is referred to as ligand-based drug design and the second, structure-based drug design.

Ligand-based design and screening

Computational approaches to drug design and screening can be either ligand- or target-based. If for a given therapeutic project, a set of active ligand molecules is known for the macromolecular target, but little or no structural information exists for the target, ligand-based computational methods can be employed. More specifically, quantitative structure activity relationship (QSAR) methods can be used, pharmacophore models developed, and shape searches performed based on the set of ligands. QSAR approaches involve the statistical analysis of a set of properties or descriptors for a series of biologically active molecules; the statistical model that is developed is then used to predict the activity of additional compounds against the target. For example, the QSAR model can be used to predict which members of a series of proposed compounds are likely to be active and therefore should be synthesized and tested. As new compounds are assayed, the additional

experimental data are used to refine the model. This general approach has been used in the pharmaceutical industry for many years to guide medicinal chemistry efforts. Typical small molecule descriptors include physicochemical properties such as molecular weight and clogP and hash codes based on the 2D structure of the ligands. Statistical approaches for determining the model may involve traditional least-squares optimization, neural networks, principal component analysis, etc. A pharmacophore model can also be derived from a set of known ligands for the target. Traditionally, a pharmacophore is the set of features common to a series of active molecules, where features can include acceptors, donors, ring centroids, hydrophobes, etc. A 3D pharmacophore specifies the spatial relationship between the groups or features, often defining distance ranges between groups, angles between groups or planes, and exclusion spheres. The programs like Phase, Catalyst and UNITY can search large 2D or 3D molecular databases for additional molecules that possess the pharmacophore. Given just one active ligand known to bind to the target, a shape search can be performed, whereby 3D molecular databases are searched for other compounds that have the same shape. Knowledge of the bound conformation of the ligand is highly desirable and again can be obtained via NMR. If the bound ligand structure is not known experimentally, the lowest energy conformation of the small molecule in solution can be calculated and used for the shape search. With certain shape search methods some chemical matching can also be specified in addition to shape fit. 3D molecular databases can then be searched for other compounds that fit into that shape.

Target-based design and screening

'Structure-based' computational approaches require the 3D structure of the target. Typically, a high-resolution (< 2.5 Å) structure from X-ray crystallography or a suitably well-defined structure from NMR spectroscopy is used. It is also possible to design ligands for and screen against a homology model for which there is a high degree of confidence. In the structure-based drug design process, once a lead compound is identified, the 3D structure of the lead compound bound to target is solved. The structure of the ligand-target complex is examined and the interactions that the small molecule makes with the target are identified. At this point, computational methods play an important role in designing modifications to existing leads. The same computational methods can also be used for the design of entirely new compounds based on the 3D structure of the target alone. Given a

lead compound, an iterative process of designing improvements to the existing lead, synthesizing and testing the designed compounds, and solving the 3D structure of each improved compound with the target begins. Many iterations of this cycle are carried out with the computational chemist/molecular modeler working closely with medicinal chemists and biologists, until a highly potent compound or series of compounds has been developed. The best compounds then enter animal trials where efficacy, toxicity, and pharmacokinetic properties are studied. Efforts to identify toxic components or potential pharmacokinetic problems such as solubility, serum albumin binding, etc, during the design phase of the project are beginning to have an impact as the necessary computational tools are improved.

Present and future

Today, computer-aided drug design and screening methods impact the efforts of all pharmaceutical companies. As the computational technologies advance, the role they play in improving the efficiency of the drug discovery process will become increasingly important. In addition, as the body of structural information on potential therapeutic targets dramatically expands, which is expected to happen in the next few years, it will drive the development of the computational methodology. Greater automation, faster algorithms, and improved information management techniques will be required to handle the sheer volume of target-related information that will need to be processed. On a genomic scale, instead of looking at individual targets, families of related targets will be studied. The information available on ligand binding to these families will be vastly expanded. The job of the molecular modeler will be to effectively mine this data as well as translate the available structural information into a form directly usable by the bench chemist. This mission will ultimately cause a greater interface of bio- and chemoinformatics, leading to improved structural and functional genomics knowledge.

Importance of protein and ligand preparation

The quality of any docking results depends on reasonable starting structures for both the protein and the ligand. It is strongly recommended that you process protein and ligand structures with these facilities in order to achieve the best results.

Protein and preparation:

A typical PDB structure file consists only of heavy atoms, can contain waters, cofactors, and metal ions, and can be multimeric. The structure generally has no information on bond orders, topologies, or formal atomic charges. Terminal amide groups can also be misaligned, because the X-ray structure analysis cannot usually distinguish between O and NH2. Ionization and tautomeric states are also generally unassigned. Most of the docking programs requires bond orders and ionization states to be properly assigned and performs better when side chains are reoriented when necessary and steric clashes are relieved.

The water molecules present in the PDB structure are identified by the oxygen atom, and usually do not have hydrogens attached. Generally, all waters (except those coordinated to metals) are deleted, but waters that bridge between the ligand and the protein are sometimes retained. If waters are kept, hydrogens will be added to them.

The atom types for metal ions are sometimes incorrectly translated into dummy atom types when metal-protein bonds are specified in the input structure. It may be necessary to adjust the protonation of the protein, which is crucial when the receptor site is a metalloprotein such as thermolysin or an MMP. Finally the protein should be minimized to reorient side-chain hydroxyl groups and alleviate potential steric clashes present in the PDB structure

Schrödinger offers a comprehensive protein preparation facility in the Protein Preparation Wizard, which is designed to ensure chemical correctness and to optimize protein

Ligand Preparation:

To give the best results, the structures that are docked must be good representations of the actual ligand structures as they would appear in a protein-ligand complex. Most of the docking tools only modify the torsional internal coordinates of the ligand during docking, so the rest of the geometric parameters must be optimized beforehand. This means that the structures supplied to docking tool must meet the following conditions:

1. They must be three-dimensional (3D)

2. They must have realistic bond lengths and bond angles.

3. They must each consist of a single molecule that has no covalent bonds to the receptor, with no accompanying fragments, such as counter ions and solvent molecules.

4. They must have all their hydrogens (filled valences).

5. They must have an appropriate protonation state for physiological pH values (around 7).

Schrödinger offers the ligand preparation facility in LigPrep. All of the above conditions can be met by using LigPrep to prepare the structures. The LigPrep process consists of a series of steps that perform conversions, apply corrections to the structures, generate variations on the structures, eliminate unwanted structures, and optimize the structures.

ADMET

Drug discovery and development are expensive and time-consuming processes. Recognition by the pharmaceutical industry that undesirable absorption, distribution, metabolism and excretion (ADME) properties of new drug candidates are the cause of many clinical phase drug development failures has resulted in a paradigm shift to identify such problems early in the drug discovery process Thus, in vitro approaches are now widely used to investigate the ADME properties of new chemical entities and, more recently, computational (in silico) modelling has been investigated as a tool to optimise selection of the most suitable drug candidates for development. The objectives of in silico modeling tools for predicting these properties to serve two key aims — first, at the design stage of new compounds and compound libraries so as to reduce the risk of late-stage attrition; and second, to optimize the screening and testing bylooking at only the most promising compounds.

Drug-like properties. The properties which can differentiate drugs from other chemicals can be considered as drug like properties. The crutial properties that should be considered for compounds with oral delivery (Lipinski's 'rule-of-five') includes molecular mass <500 daltons (Da), calculated octanol/water partition coefficient (CLOGP) <5, number of hydrogen-bond donors <5 and number of hydrogen-bond acceptors <10. These properties are then typically used to construct predictive ADME models and form the basis for what has been called property-based design.

What ADME properties do we want to predict?

A deeper understanding of the relationships between important ADME parameters and molecular structure and properties has been used to develop *in silico* models that allow the early estimation of several ADME properties. Among other important issues, we want to predict properties that provide information about dose size and dose frequency such as oral absorption, bioavailability, brain penetration, clearance (for exposure) and volume of distribution (for frequency). As a result of the availability of experimental data in the literature, considerable effort has gone into the development of models to predict physicochemical properties relevant to ADME, such as lipopshilicity. However, despite its importance, the prediction of pharmacokinetic properties such as clearance, volume of distribution and half-life directly from molecular structure is making slower progress owing to a lack of published data. Similarly, the prediction of various aspects of metabolism and toxicity is also underdeveloped.

Prediction of ADME and related properties

Absorption: For a compound crossing a membrane by purely passive diffusion, a reasonable permeability estimate can be made using single molecular properties, such as log D or hydrogen-bonding capacity. The simplest insilico models for estimating absorption are based on a single descriptor, such as log P or log D, or polar surface area, which is a descriptor of hydrogen-bonding potential. Different multivariate approaches, such as multiple linear regressions, partial least squares and artificial neural networks, have been used to develop quantitative structure–human-intestinal-absorption relationships.

Bioavailability: Important properties for determining permeability seem to be the size of the molecule, as well as its capacity to make hydrogen bonds, its overall lipophilicity and possibly its shape and flexibility.

Blood–brain barrier penetration: Drugs that act in the CNS need to cross the blood–brain barrier (BBB) to reach their molecular target. By contrast, for drugs with a peripheral target, little or no BBB penetration might be required in order to avoid CNS side effects. 'Rule-of-five'-like recommendations regarding the molecular parameters that contribute to the ability of molecules to cross the BBB have been made to aid BBB-penetration predictions; for example, molecules with a molecular mass of <450 Da or with PSA <100 Å2 are more likely to penetrate the BBB.

Dermal and ocular penetration: The existing transdermal models are typically a function of the octanol/water partition coefficient and terms that have been associated with aqueous solubility, including hydrogen-bonding parameters,molecular weight and molecular flexibility. Commercial models for the prediction of solute-permeation rates through the skin are available, for example, the QikProp and DermWin programs.

Metabolism: In silico approaches to predicting metabolism can be divided into QSAR and three-dimensional- QSAR studies, protein and pharmacophore models and predictive databases. Some of the first-generation predictive-metabolism tools currently require considerable input from a computational chemist, whereas others can be used as rapid filters for the screening of virtual libraries, for example, to test for CYP3A4 liability. Perhaps the most intellectually satisfying molecular modeling studies are those based on the crystal structure of the metabolizing enzymes several approaches that use databases to predict metabolism are available. Ultimately, such programs might be linked to computer-aided toxicity prediction on the basis of quantitative structure–toxicity relationships and expert systems for toxicity evaluation

In silico prediction of toxicity issues

Toxicity is responsible for many compounds failing to reach the market and for the withdrawal of a significant number of compounds from the market once they have been approved. It has been estimated that ~20–40% of drug failures in investigational drug development can be attributed to toxicity concerns.The existing commercially available *in silico* tools for forecasting potential toxicity issues can be roughly classified into two groups. The first approach uses expert systems that derive models on the basis of abstracting and codifying knowledge from human experts and the scientific literature. The second approach relies primarily on the generation of descriptors of chemical structure and statistical analysis of the relationships between these descriptors and the toxicological end-point.

The primary emphasis of the current software packages is carcinogenicity and mutagenicity, although some packages do also include models and/or knowledge bases for other end-points, such as teratogenicity, irritation, sensitization,immunotoxicology and neurotoxicity. There is currently an unmet need for *in silico* predictive toxicology software

for other end-points important in drug development, such as QT prolongation hepatotoxicity and phospholipidosis.

QikProp is a quick, accurate, easy-to-use absorption, distribution, metabolism, and excretion (ADME) prediction program present in the Schrödinger suite. QikProp predicts physically significant descriptors and pharmaceutically relevant properties of organic molecules, either individually or in batches. In addition to predicting molecular properties, QikProp provides ranges for comparing a particular molecule's properties with those of 95% of known drugs. QikProp also flags 30 types of reactive functional groups that may cause false positives in high-throughput screening (HTS) assays

Binding Site Analysis

Understanding the structure and function of protein binding sites is a cornerstone of structure-based drug design. Developing this understanding requires knowledge of both the location and physical properties of the binding site. In addition, the identification of small-molecule binding sites as modulators of protein-protein interactions is of increasing interest. Furthermore, even when a validated binding site has been identified, it is often important to find additional potential binding sites where appropriate targeting could result in different biological effects or new classes of compounds. When the binding site is not known from a 3-D structure or from other experimental data, computational methods can be employed to suggest likely locations. When the location of the primary binding site is known, medicinal chemistry efforts to design better ligands can profit from a better understanding of the degree to which known ligands are, or fail to be, complementary to the receptor as well as from a critical assessment of the degree to which the occupancy of accessible but unexplored regions by appropriate ligand functionality can be expected to promote binding or could be used to improve the physical properties of the ligand without lessening its binding affinity. Such assessments can assist in the evaluation and optimization both of known binding molecules and of virtual screening hits. It is also important to understand the potential druggability of the site.

SiteMap panal present in Schrödinger generates information on the character of binding sites using novel search and analysis. It generates the sites present in protein assesses each site by

calculating various properties. It also generates contour maps (site maps), producing hydrophobic and hydrophilic maps (donor, acceptor, and metal-binding regions).

Molecular Docking

Docking procedures aim to identify correct poses of ligands in the binding pocket of a protein and to predict the affinity between the ligand and the protein. In other words, docking describes a process by which two molecules fit together in three-dimensional space.

Basic Requirements for Molecular Docking

The setup for a ligand docking approach requires the following components: A target protein structure with or without a bound ligand, the molecules of interest or a database containing existing or virtual compounds for the docking process, and a computational framework that allows the implementation of the desired docking and scoring procedures. The three-dimensional structure of the protein ligand complex has to be detailed at atomic resolution. In many cases only the unbound (ligand-free, apo) form of the protein is determined, without the bioactive conformation of the ligand. Most docking algorithms assume the protein to be rigid, according to the high computational cost that the demand of flexibility implicates. The ligand is mostly regarded as flexible. Beside the conformational degrees of freedom the binding pose in the protein's binding pocket must be taken into consideration. Docking can be performed by placing rigid molecules or fragments into the protein's active site using different approaches like the clique-search, geometric hashing, or pose clustering. The flexibility of the ligand can be represented by a set of conformers covering the conformational space in an exhaustive way.

Scoring Methods

Scoring of docked poses is still regarded as one of the major challenges in the field of molecular docking. The purpose of the scoring procedure is the identification of the correct binding pose by its lowest energy value, and the ranking of protein-ligand complexes according to their binding affinities. Scoring functions can be divided in empirical scoring functions, scoring functions derived from force fields, and knowledge-based scoring functions. Scoring functions derived from force fields handle the ligand binding prediction

with the use of potential energies (non-bonded interaction terms) and sometimes in combination with solvations and entropy contributions. Knowledge-based scoring functions are based on atom pair potentials derived from structural databases. Forces and potentials are collected from known protein-ligand complexes to get a score for their binding affinities (e.g. PMF). Empirical scoring functions derive from training sets of protein-ligand complexes with determined affinity data. One general aspect in the finding of an accurate empirical scoring function is the assumption that each occurrence of an individual interaction is considered as equivalent.

Schrödinger implements docking using Glide (Grid-based Ligand Docking with Energetics). Glide searches for favorable interactions between one or more ligand molecules and a receptor molecule, usually a protein. Each ligand must be a single molecule, while the receptor may include more than one molecule, e.g., a protein and a cofactor. Glide can be run in rigid or flexible docking modes; the latter automatically generates conformations for each input ligand.

InducedFit Docking

In standard virtual docking studies, ligands are docked into the binding site of a receptor where the receptor is held rigid and the ligand is free to move. However, the assumption of a rigid receptor can give misleading results, since in reality many proteins undergo side-chain or backbone movements, or both, upon ligand binding. These changes allow the receptor to alter its binding site so that it more closely conforms to the shape and binding mode of the ligand. This is often referred to as "induced fit" and is one of the main complicating factors in structurebased drug design. The ability to model induced-fit docking has two main applications:

• Generation of an accurate complex structure for a ligand known to be active but that cannot be docked in an existing (rigid) structure of the receptor.

• Rescue of false negatives (poorly scored true binders) in virtual screening experiments, where instead of screening against a single conformation of the receptor, additional conformationsobtained with the induced fit protocol are used.

Schrödinger has developed and validated an Induced Fit Docking (IFD) protocol based on Glide and the Refinement module in Prime that accurately predicts ligand binding modes and concomitant structural changes in the receptor.

Virtual combinatorial chemistry

Recent advances in combinatorial chemistry have made it possible for chemists to synthesize large libraries of compounds, and today, highthroughput screening (HTS) allows considerable reduction of the time amount needed for the discovery of new molecules possessing biological activity for a certain target. However, the number of compounds that can be synthesized is still a small percentage of the total number of compounds that are possible in principle. The experimental efforts to carry out the biological screening of billions of compounds are still considerably high, and, therefore, computer-aided drug design approaches have emerged as a promising tool for helping medicinal chemists to decide what to synthesize. In this context, one of the major goals of computer-aided ligand discovery strategies is the identification of small subsystems from large groups of chemical compounds. Combinatorial libraries can contain several 1000–100 000 compounds and furthermore, libraries with a size of 109 or more molecules can be assembled. Up to this, the existing virtual chemistry space may contain perhaps 1060 possible molecules. The filtering of large databases or libraries of candidate compounds through the use of computational approaches based on discrimination functions that permit the selection of series of compounds to be tested for biological activity has been termed "virtual screening" (VS). There are problems encountered with the generation and screening of very large virtual libraries, however, within the next years all the necessary components for processing virtual libraries with as many as 1015 compounds will be in place. Now, by using a state-of-the-art virtual screening strategy, one should already be able to reduce the number of candidates to be examined experimentally by at least 9 orders of magnitude, thus ending up with some 1000 of compounds to be assayed for their biological activity. Moreover, the main goal of medicinal chemists is to develop compounds that do not fail in later research phases. Therefore, the prediction of side-effects and ADMET properties is crucial. Side-effects, however, are often linked to low target specificity. We propose a new approach, parallel screening, for assessing bioaffinity profiles of virtual compounds for

multiple targets in silico. Recently, with an aim toward increasing hit rates, the emphasis in computational and medicinal chemistry has shifted toward the rational design of small focused libraries that are biased toward one or several specific therapeutic targets. Focused library design can be conducted in several ways, depending on the initially available information about the therapeutic target, such as lead compound SAR, binding site interactions, etc. Whenever the 3D structure of the therapeutic target has been determined, either by X-ray crystallography or by NMR, structure-based combinatorial library design becomes an excellent tool to use in combination with other chemistry tools in the initial phase of a drug discovery program. Structure-based strategies will have the added advantage of introducing valuable information about the target structure into the combinatorial library design process and may significantly increase the hit rate in the final designed library. Not only are the compounds designed inside a specific binding site, but they are also synthetically feasible through combinatorial or parallel synthesis. This review will highlight the current computational strategies for structure-based combinatorial library design.

CombiGlide protocol present in Schrödinger employs combinatorial technology for lead identification and optimization. Though CombiGlide carries out combinatorial chemistry in silico, the results may be used either to design focused combinatorial libraries or to evaluate a large universe of compounds for one-off or small-scale automated synthesis, or even for idea generation.

Pharmacophore Modeling

The official 1998 IUPAC definition 1 is as follows: "A pharmacophore is the ensemble of steric and electronic features that is necessary to ensure the optimal supramolecular interactions with a specific biological target structure and to trigger (or to block) its biological response."

A pharmacophore does not represent a real molecule or a real association of functional groups, but a purely abstract concept that accounts for the common molecular interaction capacities of a group of compounds towards their target structure. The pharmacophore can be considered as the largest common denominator shared by a set of active molecules.

Central to the pharmacophore concept is the notion that the molecular recognition of a biological target shared by a group of compounds can be ascribed to a (small) set of common features that interact with a set of complementary sites on the biological target. In pharmacophore research quite general features such as hydrogen-bond donors, hydrogen-bond acceptors, positively and negatively charged groups, and hydrophobic regions are typically used. The other key component of contemporary pharmacophore research is the incorporation of information about the three-dimensional nature of molecular interactions. The focus of this perspective is on 3Dpharmacophore methods in which the spatial relationship between the pharmacophore features is also specified

Pharmacophore Elucidation:

Pharmacophore elucidation is a molecular alignment problem, the aim being to superimpose a set of active ligands, all of which bind to the same protein of unknown 3D structure, so that the features they have in common become evident. A number of programs for pharmacophore elucidation are widely used largely because of their availability in commercial software packages. These include CATALYST,16 GALAHAD, 17 GASP,18 the pharmacophore module of MOE,19 and PHASE.20 All pharmacophore elucidation algorithms must include methods for (a) representing the ligands (i.e., placing points on or around the molecules to represent the various pharmacophoric features they contain), (b) searching for candidate alignments, (c) scoring those alignments. These aspects are considered separately. Three main stages can be identified in the elucidation of a pharmacophore. First, prepare the data set. Second, generate possible pharmacophores. Third, validate the pharmacophore(s).

3D Database Searching

One of the common purposes of making pharmacophore models is to search for novel chemical matter. The pharmacophore represents an abstraction that can be used to find alternative chemotypes (i.e., chemical series with a different underlying framework, scaffold, or common moiety). Depending on the precision of the query, one can find numbers of hits from 10s to 1000s, which was in line with the screening capacities available at the time. Many will be false positives and show no activity in the screen, but generally, the hit rates from pharmacophore searches are much higher than from random

screening. The hits can also sample very novel and diverse chemotypes, allowing the medicinal chemist the luxury of pursuing the series with the best overall profile.

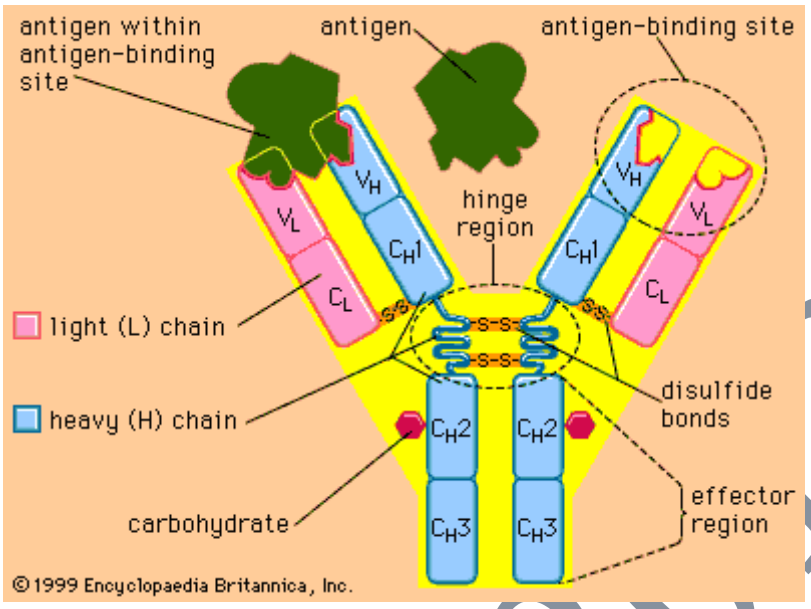
In Schrödinger, pharmacophore studies are implemented in Phase. Phase is a versatile product for pharmacophore perception, structure alignment, activity prediction, and 3D database searching. Given a set of molecules with high affinity for a particular protein target, Phase uses fine-grained conformational sampling and a range of scoring techniques to identify common pharmacophore hypotheses, which convey characteristics of 3D chemical structures that are purported to be critical for binding. Each hypothesis is accompanied by a set of aligned conformations that suggest the relative manner in which the molecules are likely to bind.

Phase consists of the following four workflows:

- *Building a pharmacophore model (and an optional QSAR models) from a set of ligands*
- *Building a pharmacophore hypothesis from a single ligand (and editing it)*
- *Preparing a 3D database that includes pharmacophore information*
- *Searching the database for matches to a pharmacophore hypothesis*

Antibody Modeling

Therapeutic antibodies are powerful tools for the prevention and treatment of human and infectious disease because of their high affinity and specificity for target antigens. Since the first U.S. Food and Drug Administration approval of a therapeutic antibody over two decades ago, antibody drugs have become available to treat cancer, infectious and cardiovascular diseases, arthritis, inflammation and immune disorders. Discovery of high-affinity antibodies for target antigens are useful in preventing many diseases more over antibodies are unique in their ability to recognize a vast diverse array of antigens, and therefore crystal structure may not be readily available for most newly developed antibody sequences, in which case a high-resolution antibody homology model is needed to perform structure-based in silico antibody engineering. Homology models can be particularly valuable if they are useful for docking or structure-based protein engineering such as increasing stability and affinity.



Antibody molecules have a common structure of four peptide chains. This structure consists of two identical **light (L) chains,** and two identical **heavy (H) chains**. Each light chain is bound to a heavy chain by a disulfide bond, and by such noncovalent interactions as salt linkages, hydrogen bonds, and hydrophobic bonds, to form a heterodimer (H-L). Similar noncovalent interactions and disulfide bridges link the two identical heavy and light (H-L) chain combinations to each other to form the basic four-chain (H-L) antibody structure, a dimer of dimers. The first 110 or so amino acids of the amino-terminal region of a light or heavy chain varies greatly among antibodies of different specificity. These segments of highly variable sequence are called *V regions:*VL in light chains and VH in heavy. All of the differences in specificity displayed by different antibodies can be traced to differences in the amino acid sequences of V regions. In fact, most of the differences among antibodies fall within areas of the V regions called *complementarity-determining regions (CDRs),* and it is these CDRs, on both light and heavy chains, that constitute the antigen binding site of the antibody molecule. By contrast, within the same antibody class, far fewer differences are seen when one compares sequences throughout the rest of the molecule.

The wide range of specificities exhibited by antibodies is a function of the sequence and length variability of six hypervariable loops or CDRs (L1, L2, L3 and H1, H2, H3), which form the antigen combining site. These six CDRs supported on a highly conserved framework region constitute the variable region of the antigen binding fragment (Fab). Since the framework region is conserved, it has proved relatively easy to model, whereas

the CDRs, by their very nature, present a more challenging problem since accuracy in their modeling is of vital. The computational approaches taken to modeling the antibody combining site, so far, fall into two groups: **knowledge based and ab initio**. In knowledge based (homology modeling) approaches, the modelling of CDRs of unknown antibodies is based on the known antibody crystal structures and select CDRs from these on the basis of length and/or sequence. The second approach has been to use **ab initio** conformational search algorithms to saturate the conformational space available to a loop and select an appropriate structure on the basis of its energy, calculated by using an empirical energy function. These methods can accurately predict thee conformations of the non-H3 CDR but difficulties in modelling the CDR H3 loop is difficult as this loop is having larger length and highly mutated.

Antibody modelling using Schrodinger:

BioLuminate serves as a gateway to a straightforward antibody-specific homology modeling workflow including automated prediction of CDR loops from sequence. BioLuminate also includes a curated antibody database with tools to add in-house antibody structures. Modeling of the Fv region of an antibody involves prediction of both the framework (FR) region and the variable loop (CDR) region. These regions are identified *automatically and their structure predicted by homology, based on known antibody structures from the PDB, or from your own database of antibodies. You can also use input coordinates for some parts of the structure and predict the rest. Finally, the H3 loop can be refined after the initial structure is generated.*

Protein-Protein Docking:

An important step towards understanding protein-protein interactions is protein-protein docking. Proteins are folded chains of amino acid polymers and together with lipids, carbohydrates and nucleic acids form the structural and functional building blocks in our cells. Flexibility of proteins makes the search for the required conformation through experimentation even more difficult. Hence, the need for fast and robust computational approaches to predicting the structures of protein-protein interactions is growing. Functions of these building blocks, and particularly those of proteins are expressed through their mutual structural interactions. The determination of the binding mode and

affinity between the constituent molecules in molecular recognition is crucial to understanding the interaction mechanisms and to designing therapeutic interventions. Due to the difficulties and economic cost of the experimental methods for determining the structures of complexes, computational methods such as molecular docking are desired for predicting putative binding modes and affinities. There are two main aspects of a docking algorithm; scoring or measuring the quality of any given docked complex, and searching for the highest scoring or a pool of high quality docking conformations.

The goal of protein–protein docking is to determine the structure of a complex in atomic detail, starting from the coordinates of the unbound component molecules. Most of the current docking methods start with rigid body docking that generates a large number of docked conformations with good surface complementarity. The Fast Fourier Transform (FFT) correlation approach, introduced in 1992 by Katchalski-Katzir and coworkers, revolutionized this step of rigid body search. Owing to the numerical efficiency of this algorithm it became computationally feasible, for the first time, to systematically explore the conformational space of protein–protein complexes evaluating the energies for billions of conformations on a grid, and thus to dock proteins without any a priori information on the expected structure. Other approaches, primarily Monte Carlo, also perform well if the search can be restricted to regions of the conformational space, but become computationally expensive if no such constraints are available. For this reason, FFT-based docking is the first step in many methods that have performed well at CAPRI (Critical Assessment of Predicted Interactions), the first community-wide experiment devoted to protein docking.

Protein-Protein Docking with Schrodinger BioLuminate

Protein-protein docking in BioLuminate is performed using the Piper program, under license from Boston University. The job can be set up in the Protein-Protein Docking panel. In this panel you can set up jobs to dock two arbitrary proteins, dock an antigen to an antibody, or dock one protein to itself to form a dimer or a trimer.One protein is treated as the "receptor" and the other as the "ligand". In the general case, it does not matter which protein is treated as the receptor and which protein is treated as the ligand. For antibody-antigen docking, the receptor is the antibody and the ligand is the antigen.The algorithm

samples all possible orientations of the two proteins, subject to whatever constraints are applied. It uses a grid to locate the best poses of the two proteins, with a maximum resolution in the poses of about 5°. The docking is performed as a rigid-body optimization: there is no subsequent minimization of the interfacial region.

Residue Scanning and Hot spot prediction:

Protein-protein interactions are key components of all signal transduction processes, mediating the integration of linear pathways into the often complex interaction networks revealed by recent genome-scale studies. Tools to rationally alter and interfere with protein interactions offer great promise to help dissect the function of connectivities in these networks. The ability to alter protein interactions requires an understanding of the determinants of affinity and specificity in protein interfaces. Experimental "alanine-scanning mutagenesis" is a powerful method for analyzing important interactions in protein-protein interfaces.Alanine scanning measures the effect of the deletion of an amino acid side chain beyond the C beta carbon atom on the affinity of a protein-protein complex.

Experimentally, a hot spot can be found by evaluating free energy change upon mutating it to an alanine, playing key roles on the stability of the protein association. Binding Interface Database (BID) (Fischer et al., 2003) presents experimentally verified hot spots at interfaces collected from literature.

Analysis of amino acid composition of hot spots shows that some residues are more favorable. The most frequent ones, Tyr,Arg and Trp, are critical due to their size and conformation in hot spots. Hot spot information from experimental studies are available only for a very limited number of complexes, therefore, there is a need for computational methods to identify hot spots of protein interaction sites (DeLano, 2002). Protein interfaces accounting for energies of packing interactions, hydrogen bonds and solvation. Computational hot spots, the residues they identified computationally based on their model, show accordance with experimental hot spots in ASEdb.

Some examples of the use of residue scanning are:

• improvement of protein-ligand affinity

• identifying protein-protein interface hotspots

• identifying residue mutations that can improve stability.

• mutating unpaired, solvent-exposed Cys residues to reduce undesired reactivity

Protein engineering with Schrodinger BioLuminate:

BioLuminate can be used to visualize protein aggregation propensity surfaces and the results of ***residue-based property predictions*** *including predictions of binding energy, thermal stability, solvent-accessible surface area, hydrophilicity, and hydrophobicity. In addition, cysteine scanning results and reactive hot spots can be visualized.*

Glossary

Quantum Mechanics:

***Ab Initio* calculations**: *Ab Initio* (from the beginning) calculations are the computations which are directly derived from theoretical principles (such as the Schrödinger equation), with no inclusion of experimental data.

Semi empirical Method: Uses simplifications of the Schrödinger equation H Y = E Y to estimate the energy of a system (molecule) as a function of the geometry and electron distribution.The simplifications require empirically derived (not theoretical) parameters (or fudge factors) to allow calculated values to agree with observed values.

Molecular Mechanics: Applicable only to parameterized systems.Molecule is described as a series of charged points (atoms) linked by springs (bonds).Connectivity of atoms cannot change during the simulation (no chemical reactions).Can simulate behavior of systems with 1000's of unique atoms

Force Fields: A force field (also called a forcefield) refers to the functional form and parameter sets used to describe the potential energy of a system of particles (typically but not necessarily atoms). Force field functions and parameter sets are derived from both experimental work and high-level quantum mechanical calculations. "All-atom" force fields provide parameters for every atom in a system, including hydrogen, while "united-atom"

force fields treat the hydrogen and carbon atoms in methyl and methylene groups as a single interaction center. "Coarse-grained" force fields, which are frequently used in long-time simulations of proteins, provide even more abstracted representations for increased computational efficiency.

Energy Minimization: The process by which the potential energy of a molecule is brought to its closest local minimum is known as minimization.

Global Minima: The minimum point on the Potential Energy Surface (P.E.S) with very lowest energy is known as the Global Energy minimum. To find lowest P.E. the molecular structure (defined in terms of atom positions) is varied hereby producing a change in bond lengths, angles etc. as well as in the conformation.

Local Minima: Minimum energy points which correspond to stable structures on Potential Energy Surface are referred to as Local minima

Saddle Point: The highest point on the pathway between 2 minima is known as the saddle point with the arrangement of the atoms being the transition structure

Conformational Search: Exploring the conformations of a molecule by rotating single bonds. The conformational search identifies the "preferred" (low energy) conformations of a molecule **Molecular Dynamics**: Molecular Dynamics is a deterministic process based on the simulation of molecular motion by solving Newton's equations of motion for each atom and incrementing the position and velocity of each atom by use of small time increment.

Ensemble: An ensemble is a collection of all possible systems which have different microscopic states but have an identical macroscopic or thermodynamic state.

Statistical Mechanics: Statistical Mechanics is the mathematical means to extrapolate the thermodynamic properties of bulk materials from a molecular description of the material

Homology Modeling: Homology modeling, also known as comparative modeling of protein refers to constructing an atomic-resolution model of the "target" protein from its amino acid sequence and an experimental three-dimensional structure of a related homologous protein (the "template").

Ligand Pose: The combination of position and orientation of a ligand relative to the receptor, along with its conformation in flexible docking, is referred to as a *ligand pose.*

Contact us

Dr. J. Jeyakanthan
Convenor, SBCADD'2013

Dr. P. Srinivasan

Organizing Secretary, SBCADD'2013
Department of Bioinformatics
Science Block, 4th Floor
Alagappa University
Karaikudi - 630 004
Tamil Nadu, India.
Mobile: +91-9444482814

+ 91-4565-230725
+91- 4565-225202
dbiconf@gmail.com
www.alagappauniversity.ac.in and
www.bioinfoau.org