

Review Article

AN INSIGHT TO VIRTUAL LIGAND SCREENING METHODS FOR STRUCTURE-BASED DRUG DESIGN AND METHODS TO PREDICT PROTEIN STRUCTURE AND FUNCTION IN LUNG CANCER: APPROACHES AND PROGRESS

BHAGAVATHI S*, PRAKASH A, GULSHAN WADHWA

Department of Biotechnology, Bioinformatics Centre Barkatullah University, Bhopal.
Email: bhagavathikanagaraj@gmail.com

Received: 29 Jul 2014 Revised and Accepted: 14 Sep 2014

ABSTRACT

Lung cancer is a complex disease that involves multiple types of biological interactions across diverse, physical, temporal, and biological scales. This complexity presents substantial challenges for the characterization of lung cancer, and motivates the study of cancer in the context of molecular, cellular, and physiological systems. Computational models of Lung cancer are being developed to aid both biological discovery and clinical medicine. The development of these *in silico* models is facilitated by rapidly advancing experimental and analytical tools that generate information-rich, high-throughput biological data. Protein structure prediction by using bioinformatics can involve sequence similarity searches, multiple sequence alignments, identification and characterization of domains, secondary structure prediction, solvent accessibility prediction, automatic protein fold recognition, constructing three-dimensional models to atomic detail, and model validation. Till today technologies like combinatorial chemistry and high-throughput screening (HTS) authorize biological assays of a large number of small molecules against the therapeutically relevant targets. However, the escalating costs highlight the need of developing novel approaches while still allowing one to explore larger chemical diversity. In this respect, Virtual ligand screening (VLS) is established as an attractive approach to handle large sets of compounds and to improve the “hit-rate” of drug discovery programs. Here, we review the main Ligand Screening techniques applied for structure-based drug design and we focus on key concepts in the molecular docking–scoring methodology in lung cancer. These methods, if used appropriately, can provide valuable indicators of protein structure and function for the different type of cancers.

Keywords: Lung cancer, Molecular modeling, Sequence similarity searches, Multiple sequence alignment, Secondary structure prediction, Virtual screening, Structure-based drug design, Polo like kinase 1, Thrombomodulin, Review.

INTRODUCTION

Computational methods that predict the structure and specificity of protein-protein interactions can yield deep insight into the structural biology of many biochemical pathways. Through high-resolution structures of protein-interactions we can identify the structural mechanisms of diseases, engineer proteins towards specific functions, and design drugs that disrupt pathogenesis. Challenges in accurately modeling protein interactions include efficiently sampling the conformational space available for two proteins to interact, and adequately approximating the free energy of the conformational landscape to correctly predict the structure and specificity of the protein interaction.

Systems biology is a rapidly growing discipline that employs an integrative approach to characterize biological systems, in which interactions among all components in a system are described mathematically to establish a computable model. These *in silico* models – which complement traditional *in vivo* animal models – can be simulated to quantitatively study the emergent behavior of a system of interacting components. The advent of high-throughput experimental tools has allowed for the simultaneous measurement of thousands of biomolecules, paving the way for *in silico* model construction of increasingly large and diverse biological systems. Integrating heterogeneous dynamic data into quantitative predictive models holds great promise to significantly increase our ability to understand and rationally intervene in disease-perturbed biological systems.

This promise – particularly with regards to personalized medicine and medical intervention – has motivated the development of new methods for systems analysis of human biology and disease. Even though cancer has been among the most-studied human diseases using systems approaches, significant challenges remain before the enormous potential of *in silico* cancer biology can be fully realized.

Cancer is an intrinsically complex and heterogeneous disease, making it particularly amenable to systems biology approaches.

Malignant tumors develop as a function of multiple biological interactions and events, both in the molecular domain among individual genes and proteins, and the cellular and physiological levels between functionally diverse somatic cells and tissues [1]. (fig.1)

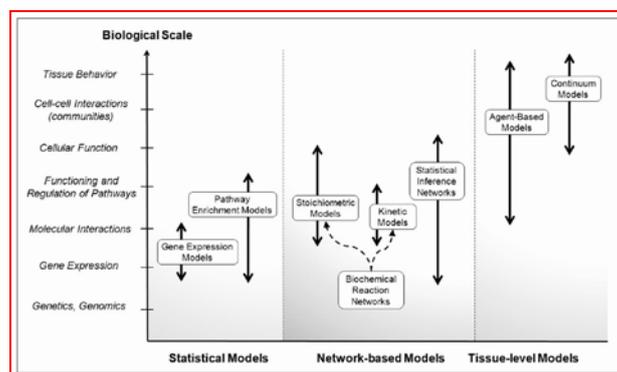


Fig. 1: Biological scales and potential modeling approaches

At the molecular level, genetic lesions interact synergistically to evade tumor suppression pathways, with not even a single mutation typically sufficient to cause transformation [2–6]. Beyond genetic mutations, transformed cells can exhibit changes in expression of hundreds to thousands of genes and proteins [7–9]. Genetic modifications observed in cancer are often accompanied by changes at the epigenetic level [10–15].

The convolution of genetic effects and epigenetic modifications illustrates the complex, nonlinear relationship between molecular state and cellular cancer phenotype, emphasizing the need for heterogeneous data integration through *in silico* models. (fig.2)

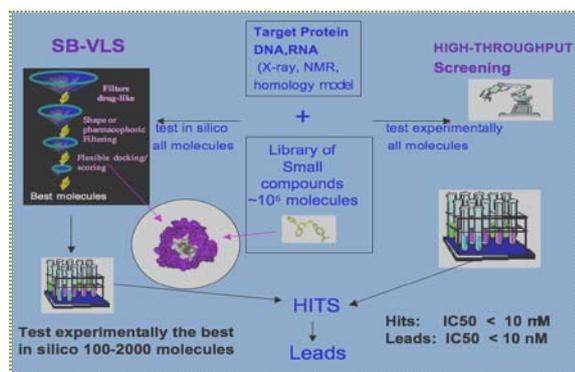


Fig. 2: Flowchart of Structure-based Virtual Ligand Screening (SB-VLS) versus High-Throughput Screening (HTS) for hit identification.

Virtual Ligand Screening: Another alternative to HTS (High throughput screening)

The current post genomic era has been characterized by a large increase in the number of potential therapeutic targets amenable to investigation. This growth is, in turn, increasing pressure on the ability of the pharmaceutical industry to prioritize programmes and conduct lead discovery in a highly efficient manner. Such pressure is evident in a recent statistic: Over the past 10 years only 25% of quality targets have yielded a quality lead series. Logistical reasons could be put forward to explain this, such as inefficient programme management, but the overriding significance of this figure is that the technology used to generate leads from a given target is not functioning efficiently. The financial implications of this high attrition rate, even to the lead identification stage, are huge. There is an urgent need therefore to review the technologies currently employed in lead identification and critically assess which methodologies are likely to increase productivity at the early discovery stages. High Throughput Screening (HTS) has traditionally been the most widely used methodology at the hit finding stage of the drug discovery process. In contrast to HTS, lead discovery using a target structure as a starting point for computational techniques to screen. Design and prioritise compounds is promising to be a much more efficient process. This *in silico* structure-based design is rapidly becoming the lead identification cornerstone of many drug discovery processes. VS has quickly gained popularity because it can be used to screen a company's current compound collection and hits resulting from this process can be tested in a time efficient manner (without synthesis usually). This is the simplest use case of VS and reflects its main application today. Regardless of research, there is always likely to be a trade-off between accuracy of scoring and the genuine high throughput nature of Virtual Screening.

Virtual ligand screening based on the 3D structure of macro molecular targets (structure-based SB-VLS) is widely applied to identify chemical entities that have a high likelihood of binding to a target molecule to elicit desired biological responses [16-19]. For SB-VLS methods (FIG. 3), it is assumed that the 3D structure of the target is known either by X-ray crystallography or NMR experiments, or predicted by homology modeling [20-23]. The principle here is to dock all the ligands present in a database into the binding pocket. In principle, this approach accesses a vast virtual chemistry space, far in excess of what could ever be biologically screened. As such, this process is likely to locate many new scaffolds, critical in 'me too' drug discovery programmes where the IP coverage on a given target is likely to be heavy. The downside of the vs approach is ensuring the chemical feasibility of the predicted structures which represents an on-going technical challenge with the selected target and evaluates the fit between the molecules [24].

Bioinformatics and drug discovery

Drug discovery is the step-by-step process by which new candidate drugs are discovered. Traditionally, pharmaceutical companies follow well-established pharmacology and chemistry-based drug

discovery approaches, and face various difficulties in finding new drugs [25]. In the highly competitive "winner takes all" pharmaceutical industry, the first company to patent a new chemical entity (NCE i. e., new drug candidate) for a specific treatment takes all the spoils, leaving other competitors to mostly wait for patent expirations to partake in the largesse. Nowadays, therefore, pharmaceutical companies invest heavily in all those approaches that show potential to accelerate any phase of the drug development process [26]. The increasing pressure to generate more and more drugs in a short period of time with low risk has resulted in remarkable interest in bioinformatics [27]. In fact, now there is an existence of new, separate field, known as computer-aided drug design (CADD) [28-29].

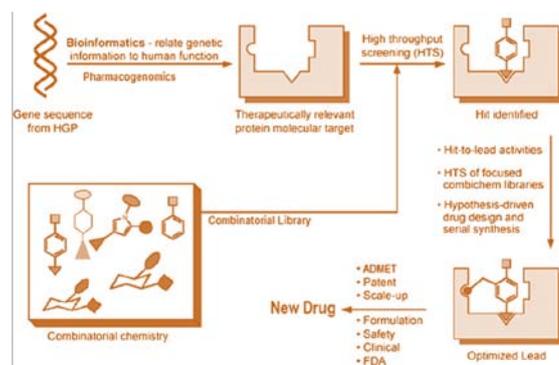


Fig. 3: The role of bioinformatics in various stages of drug discovery process

One of the major thrusts of current bioinformatics approaches is the prediction and identification of biologically active candidates, and mining and storage of related information. Drugs are usually only developed when the particular drug target for those drugs' actions have been identified and studied. The number of potential targets for drug discovery process is increasing exponentially. In addition to identifying new targets that offer more potential for new drugs [30-31]. This is an area where the human genome information is expected to play a master role [32]. Drug developers are presented with an unaccustomed luxury of choice as more genes are identified and the drug discovery cycle becomes more data-intensive [33].

Comparative homology modeling

The aim of comparative or homology protein structure modeling is to build a three-dimensional (3D) model for a protein of unknown structure (the target) on the basis of sequence similarity to proteins of known structure (the templates) [34-38]. Two conditions must be met to build a useful model. First, the similarity between the target sequence and the template structure must be detectable. Second, a substantially correct alignment between the target sequence and the template structures must be calculated. Comparative modeling is possible because small changes in the protein sequence usually result in small changes in its 3D structure [39]. Although considerable progress has been made in ab initio protein structure prediction [40], comparative protein structure modeling remains the most accurate prediction method. The overall accuracy of comparative models spans a wide range, from low resolution models with only a correct fold to more accurate models comparable to medium resolution structures determined by crystallography or nuclear magnetic resonance (NMR) spectroscopy. The 3D structures of proteins in a family are more conserved than their sequences [41]. Therefore, if similarity between two proteins is detectable at the sequence level, structural similarity can usually be assumed. Moreover, even proteins that have non detectable sequence similarity can have similar structures. It has been estimated that approximately one third of all sequences are recognizably related to at least one known protein structure [42-46]. Because the number of known protein sequences is approximately 500,000 [47], comparative modeling could in principle be applied to more than 150,000 proteins. This number can be compared to approximately

10,000 protein structures determined by experiment [48-49]. The usefulness of comparative modeling is steadily increasing because the number of unique structural folds that proteins adopt is limited [50] and because the number of experimentally determined new structures is increasing exponentially [51]. It is possible that in less than 10 years at least one example of most structural folds will be known, making comparative modeling applicable to most protein sequences. All current comparative modeling methods consist of four sequential steps (fig. 4): fold assignment and template selection, template–target alignment, model building, and model evaluation. If the model is not satisfactory, template selection, alignment, and model building can be repeated until a satisfactory model is obtained.

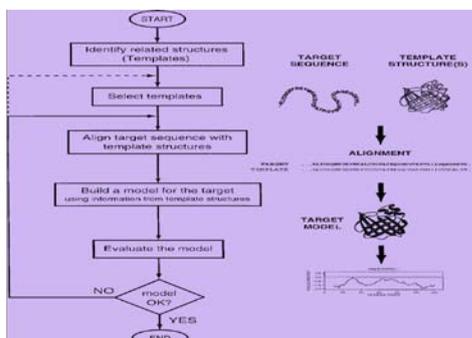


Fig. 4: Steps in comparative protein structure modeling

An introduction to docking

The docking process involves the prediction of ligand conformation and orientation (or posing) within a targeted binding site. In general, there are two aims of docking studies: accurate structural modeling and correct prediction of activity. However, the identification of molecular features that are responsible for specific biological recognition, or the prediction of compound modifications that improve potency, are complex issues that are often difficult to understand and even more so to simulate on a computer. In view of these challenges, docking is generally devised as a multi-step process in which each step introduces one or more additional degrees of complexity [52]. The process begins with the application of docking algorithms that pose small molecules in the active site. This in itself is challenging, as even relatively simple organic molecules can contain many conformational degrees of freedom. Sampling these degrees of freedom must be performed with sufficient accuracy to identify the conformation that best matches the receptor structure, and must be fast enough to permit the evaluation of thousands of compounds in a given docking run. Relatively simple scoring functions continue to be heavily used, at least during the early stages of docking simulations. Pre-selected conformers are often further evaluated using more complex scoring schemes with more detailed treatment of electrostatic and van der Waals interactions, and inclusion of at least some solvation or

entropic effects [53]. It should also be noted that ligand-binding events are driven by a combination of enthalpic and entropic effects, and that either entropy or enthalpy can dominate specific interactions. This often presents a conceptual problem for contemporary scoring functions because most of them are much more focused on capturing energetic than entropic effects. In addition to problems associated with scoring of compound conformations, other complications exist that make it challenging to accurately predict binding conformations and compound activity. These include, among others, limited resolution of crystallographic targets, inherent flexibility, induced fit or other conformational changes that occur on binding, and the participation of water molecules in protein–ligand interactions. Without doubt, the docking process is scientifically complex.

Molecular representations for docking

To evaluate various docking methods, it is important to consider how the protein and ligand are represented. There are three basic representations of the receptor: atomic, surface and grid [54]. Among these, atomic representation is generally only used in conjunction with a potential energy function [55] and often only during final RANKING procedures (because of the computational complexity of evaluating pair-wise atomic interactions). Surface-based docking programs are typically, but not exclusively, used in protein–protein docking [56-57]. Connolly's early work on molecular surface representations is mainly responsible for spawning much of the research in this area [58-59]. These methods attempt to align points on surfaces by minimizing the angle between the surfaces of opposing molecules [60].

Therefore, a rigid body approximation is still the standard for many protein–protein docking techniques. The use of potential energy grids was pioneered by Goodford [61], and various docking programs use such grid representations for energy calculations. The basic idea is to store information about the receptor's energetic contributions on grid points so that it only needs to be read during ligand scoring. In the most basic form, grid points store two types of potentials: electrostatic and van der Waals. A representative grid for capturing electrostatic potentials, and illustrates the electrostatic potential of a bound inhibitor mapped on its molecular surface.

Computing Systems for Protein Structure Prediction (review)

Rasmol [62] is a macromolecule viewer; the correct mime-types and helper-applications need to be set in the browser's preferences to view structures. Rasmol was not designed to manipulate atomic stereochemistry. Software such as Composer [63-64], Modeller [65], What If [66], SwissModel [67], and Naomi [68] are of value in protein structure modeling to atomic detail (table 1). In addition to building protein models, software packages are required to interactively visualize and monitor the building process. Commercial packages for molecular modeling, developed by Accelrys and Tripos, can provide these facilities (table 1).

These commercial products have extensive graphical user interfaces and have been developed with emphasis on ease of use and project management and continuity.

Table 1: Tools for comparative molecular modeling of protein structures

Server	Ref.	URL
Academic Versions		
COMPOSER [^]	(63,64)	http://www-cryst.bioc.cam.ac.uk/
Modeller [*]	(65)	http://guitar.rockefeller.edu/modeller/
WhatIF [^]	(66)	http://www.sander.embl-heidelberg.de/whatif/
SwissModel [^]	(67)	http://www.expasy.ch/swissmod/SWISSMODEL.html
NAOMI [^]	(68)	http://www.cambridgeantibody.com/
Commercial		
Modeller [*]	(65)	http://www.accelrys.com/
Homology [^]	(80)	http://www.accelrys.com/
DISCOVER	-	http://www.accelrys.com/
SYBYL	-	http://www.tripos.com/
COMPOSER [^]	(63,64)	http://www.tripos.com/

The key to the symbols used follows: ^{*} restraint-based molecular modeling techniques; [^] rigid-body fragment assembly techniques.

Proteins of our Interest in lung cancer

Lung cancer constitutes one of the leading causes of death in industrialized countries, and its incidence is rapidly growing in developing nations worldwide. Although tobacco smoke and other environmental pollutants are responsible for more than 80–90% of the cases in men [69] it is well established that less than 10–15% of smokers develop lung cancer, indicating that other factors might contribute to the development of this disease [70-71]. Human lung cancer cells have been found to express varying degrees of several kinds of onco developmental antigens, such as carcino embryonic antigen and stage-specific embryonic antigen related antigens, which are found expressed in stage-specific lung buds of human embryos and may play some role in the cell-to-cell interactions. Here we discuss mainly about two proteins which are highly expressed in lung cancer namely Polo like kinase1, Thrombomodulin. Thrombomodulin acts as an important oncodevelopmental antigen which is found expressed in lung cancer cells. Extensive studies have shown that Plk1 expression is elevated in non-small-cell lung cancer, head and neck cancer, esophageal cancer, gastric cancer, melanomas, breast cancer, ovarian cancer, endometrial cancer, colorectal cancer, gliomas, and thyroid cancer. Polo like kinase 1 (**Plk1**) gene and protein expression has been proposed as a new prognostic marker for many types of malignancies, and Plk1 is a potential target for cancer therapy [72-74]. Selection of a potential target for therapy is a daunting task. In-silico modeling is a multidisciplinary method integrating mathematical models with experimental (in vitro and in vivo) and clinical data [75]. Homology or evolutionary relatedness represents a key concept in studying protein sequence, structure, and function. Homologs can be inferred by sequence similarity search tools such as the popular sequence-profile comparison method PSI-BLAST [76]. Basic Local Alignment Search Tool (BLAST) provides an "expect" value, statistical information about the significance of each alignment [77]. MACS (multiple alignments of complete sequences) are typically used to perform comparative analysis at the genome level, to define the phylogenetic relationships between organisms in evolutionary studies, to identify conserved functional residues, motifs or domains and to predict protein [78]. Comparative, or homology, modelling structures is the most widely used prediction method when the target protein has homologues of known structure [79].

This study is aimed at modeling and evaluating the structure and we review the main VLS techniques applied for structure-based drug design and we focus on key concepts in the molecular docking-scoring methodology. These methods, if used appropriately, can provide valuable indicators of protein structure and function for different type of cancers.

Methods

Bioinformatics is a rapidly evolving science, and new and improved versions of the software and databanks are released frequently. Therefore, the methods are presented as a guide to the principles relevant and applicable in the field. Before setting out, it is important to have some background understanding of the following: the databank search algorithms; the information content of the databanks; sequence retrieval from the databanks; sequence alignments; protein structures and Unix commands.

Search for a structural homolog using standard search methods

A sequence similarity search can be performed to query a protein sequence against the amino-acid sequences of known 3D protein structures. If a structural homolog has been reliably identified for a significant fraction of the query sequence, a model can be built by using standard homology modeling methods [63–68] [80]. BLASTP [81] was used to detect sequence similarities in a databank of protein sequences with solved structures. The proteins of our interest are Polo like kinase 1, thrombomodulin derived from the gene expression data of Lung cancer through micro array technology.

Sequence retrieval from swissprot

Amino acid sequences retrieved from swissprot/uniprot (www.uniprot.org) provides descriptions of a non-redundant set of proteins including their function, domain structure, posttranslational modifications and variants [82] [83]. This database merges all proteins in single entry coded by one gene so as to minimize redundancy and improve reliability with fully featured information. Cross-references with other databases modemize swissprot entries to hold detailed expertise [84]. The accession number of our retrieved sequences is Q58A51 and P07204. (fig. 5 & 6).

Q58A51 (Q58A51_HUMAN) ★ Unreviewed, UniProtKB/TrEMBL
 .last modified April 18, 2012. Version 61. [History...](#)

Contribute
[Send feedback](#)
[Read comments \(0\) or add your own](#)

Clusters with 100%, 90%, 50% identity | [Third-party data](#)

[text](#) [xml](#) [rdf/xml](#) [gff](#) [fasta](#)

[Names](#) · [Attributes](#) · [General annotation](#) · [Ontologies](#) · [Sequences](#) · [References](#) · [Cross-refs](#) · [Entry info](#) · [Customize order](#)

Names and origin

Protein names	Submitted name: Polo-like kinase 1 delta EMBL BAD93352.1 Submitted name: Serine/threonine-protein kinase PLK1 Ensembl ENSP00000331713
Gene names	Name: Plk-1 delta EMBL BAD93352.1 Synonyms: PLK1 Ensembl ENSP00000331713
Organism	Homo sapiens (Human) EMBL BAD93352.1
Taxonomic identifier	9606 [NCBI]
Taxonomic lineage	Eukaryota › Metazoa › Chordata › Craniata › Vertebrata › Euteleostomi › Mammalia › Eutheria › Euarchontoglires › Primates › Haplorrhini › Catarrhini › Hominoidea › Hominidae › Homo

Fig. 5: Sequence retrieval from Swissprot page PLK-1

Protein structure optimization, quality assessment and visualization

Structural homologous entries were obtained for proteins through local alignment search using BlastP (Basic Local Alignment Search Tool) [85], against Protein Data Bank (PDB) [81].

(fig. 7 & 8) Comparison of homology models with known structure (Template) may also reveal similarities which allow biochemical and biological functions to be inferred. The alignment was used for comparative modelling to build 3D model by satisfaction of spatial

restraints using Modeller9v7 [86]. The core modelling procedure begins with an alignment of the sequence to be modelled (Target) with related known 3D structures (templates). This alignment is

usually input to the program. The output is a 3D model for the target sequence containing all main chain and side chain non hydrogen atoms.

P07204 (TRBM_HUMAN) ★ Reviewed, UniProtKB/Swiss-Prot
 Last modified April 18, 2012. Version 165. [History...](#)

Contribute
[Send feedback](#)
[Read comments \(0\) or add your own](#)

Clusters with 100%, 90%, 50% identity | Documents (7) | Third-party data
[text](#) [xml](#) [rdf/xml](#) [gff](#) [fasta](#)

Names · Attributes · General annotation · Ontologies · Interactions · Sequence annotation · Sequences · References · Web links · Cross-refs ·
 Entry info · Documents · Customize order

Names and origin

Protein names	Recommended name: Thrombomodulin Short name=TM Alternative name(s): Fetomodulin CD_antigen=CD141
Gene names	Name: THBD Synonyms: THRM
Organism	Homo sapiens (Human)
Taxonomic identifier	9606 [NCBI]
Taxonomic lineage	Eukaryota › Metazoa › Chordata › Craniata › Vertebrata › Euteleostomi › Mammalia › Eutheria › Euarchontoglires › Primates › Haplorrhini › Catarrhini › Hominidae › Homo

Fig. 6: Sequence retrieval from swissprot page Thrombomodulin

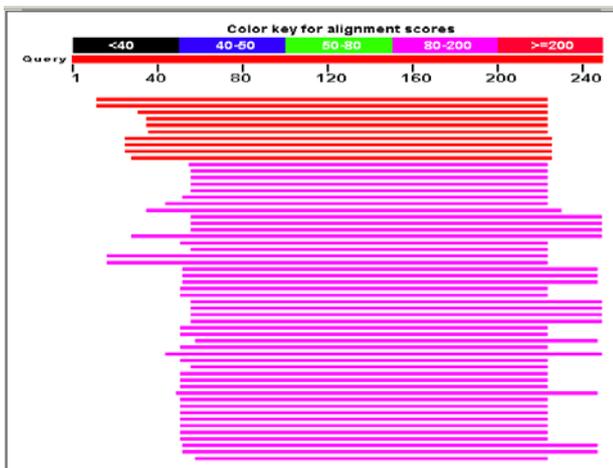


Fig. 7: Blast search and alignment for polo like kinase 1.

[pdb|3KB7|A](#) Chain A, Crystal Structure Of Polo-Like Kinase 1 In Complex With A Pyrazoloquinazoline Inhibitor
 Length=311

Score = 384 bits (987), Expect = 4e-134, Method: Compositional matrix adjust.
 Identities = 186/187 (99%), Positives = 186/187 (99%), Gaps = 0/187 (0%)

Query 36 PAKEIPEVLVDPSSRRRYVGRFLGKGGFAKCFEISDADTKEVFAGKIVPKSLLKPHQR 95
 PAKEIPEVLVDPSSRRRYVGRFLGKGGFAKCFEISDADTKEVFAGKIVPKSLLKPHQR
 Sbjct 2 PAKEIPEVLVDPSSRRRYVGRFLGKGGFAKCFEISDADTKEVFAGKIVPKSLLKPHQR 61

Query 96 ERMSMEISIHRSLAHQHVGFHGFEDNDFFVFWLELCRRRSLELHKRRKALTEPEARY 155
 ERMSMEISIHRSLAHQHVGFHGFEDNDFFVFWLELCRRRSLELHKRRKALTEPEARY
 Sbjct 62 ERMSMEISIHRSLAHQHVGFHGFEDNDFFVFWLELCRRRSLELHKRRKALTEPEARY 121

Query 156 YLRQIVLGCQYLHRRNVIHRDLKLNLFMEDLEVKIGDFGLATKVEYDGERKKTLOGTP 215
 YLRQIVLGCQYLHRRNVIHRDLKLNLFMEDLEVKIGDFGLATKVEYDGERKKTLOGTP
 Sbjct 122 YLRQIVLGCQYLHRRNVIHRDLKLNLFMEDLEVKIGDFGLATKVEYDGERKKTLOGTP 181

Query 216 NYIAPAV 222
 NYIAP V
 Sbjct 182 NYIAPEV 188

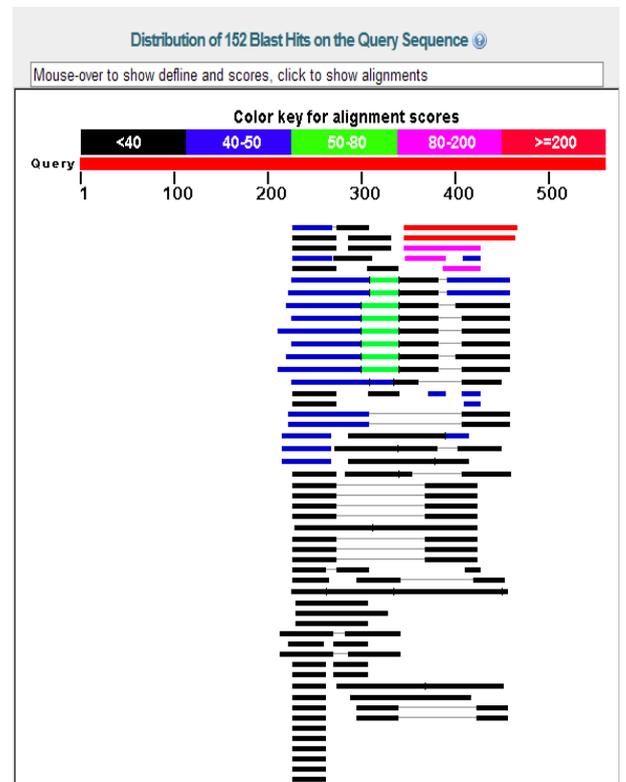


Fig. 8: Blast search and alignment for thrombomodulin

Validation of protein structure models

MODELLER generated several preliminary models which were ranked based on their DOPE scores. Some models having low DOPE score were selected and stereochemical property of each models was assessed by PROCHECK [87-88]. PROCHECK server was used for the validation of modeled Esx homeobox-1 protein structure

analysis of the model was done to check whether the residues are falling in the most favored region in the Ramachandran's plot or not. The model with the least number of residues in the disallowed region was selected for further studies. Quality of models was evaluated with respect to energy and stereochemical geometry. ProSA-Web server [89] was employed to evaluate energy and verify 3Dserver [90] to evaluate the local compatibility of the model related to good protein structure.

```

>|pdb|3P5B|S Chain L, The Structure Of The LdlrFCSK9 COMPLEX REVEALS THE RECEPTOR
IN AN Extended Conformation
Length=400

Sort alignments for this subject sequence by:
E value Score Percent identity
Query start position Subject start position

Score = 56.2 bits (134), Expect = 2e-08, Method: Compositional matrix adjust.
Identities = 28/66 (42%), Positives = 35/66 (53%), Gaps = 3/66 (5%)

Query 274 CEHFCVFNPDQSGYSVCMCEYGRVLAADQHRCEVDVDCILEPSPCPQRCVNTQGGFECHC 333
          C H C D Y C+C G+L A Q RCEB+D+C +F C Q CVH +GG+ C C
Sbjct 10 CSHVC--VDLKITGYECLCPDGFQLVA-QRCEIDEC-QDPTCSQLCVNLEGGYKQC 64

Query 334 YPHYDL 339
          + L
Sbjct 65 EEBFQL 70

Score = 45.1 bits (105), Expect = 8e-05, Method: Compositional matrix adjust.
Identities = 29/63 (35%), Positives = 35/63 (42%), Gaps = 4/63 (5%)

Query 226 DCSVENGSGCEHACNAPGAPRQCPAGALQADGRSCTASATQSCNDLCEHFCVFNPDQP 285
          +C MGGC H CN + C CP G L A R Q D C CV +
Sbjct 2 ECLDNGGCSHVNDLKITGYECLCPDGFQLVAQRCEIDECQDPTCSQLCVNLEGGYKQC 57

Query 286 GYSVCMCEYGRVLAADQHRCEVDV 308
          G Y C CE G+L C+ V
Sbjct 58 GGYKQCEEGFQLDPHTKACNIV 80

Score = 40.0 bits (92), Expect = 0.003, Method: Compositional matrix adjust.
Identities = 25/63 (37%), Positives = 32/63 (47%), Gaps = 1/63 (1%)

Query 391 NQTACPADC-DPNTQASCECEPEGYILDGFICTDIDECENGSGFCVGNLPGTFEFCIG 449
          N C C D C CP+G+ L C DIDE C+ CS +C NL G +C C
Sbjct 6 NNGGCSHVNDLKITGYECLCPDGFQLVAQRCEIDECQDPTCSQLCVNLEGGYKQC 65

Query 450 PDSALARH 457
          L H
Sbjct 66 EGFQLDPH 73
    
```

Polo like kinase 1 (Q58A51) was subjected to homology search against PDB database using BlastP to identify significant structural homologs to be used as template for homology modelling. The results indicated the presence of Pkc like superfamily domain and the best homolog was 3KB7 with 99 % identity with the query protein and thus served as a template for modelling and the modelled protein obtained is shown in (fig. 9) and validation were done using the Ramachandran map (fig.10.) After loop refinement.

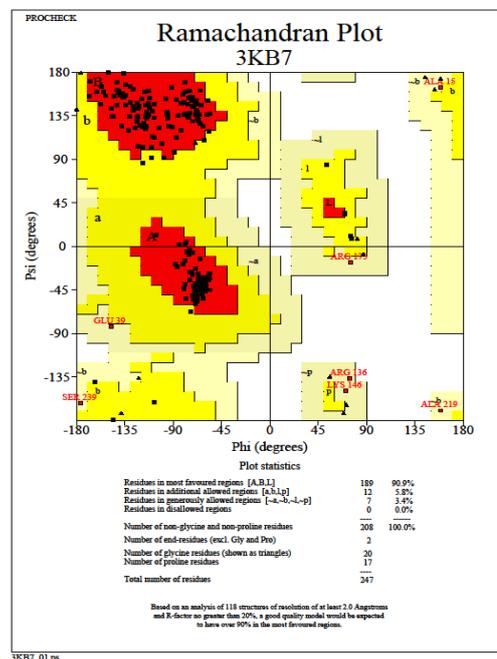


Fig.10: Ramachandran plot of Polo like kinase1

Ramachandran Analysis was performed to determine the stability of the modelled structure. Subsequently the model structure was validated using PROCHECK, which determine stereo chemical aspects along with main chain and side chain parameters with comprehensive analysis. The model obtained after refining the protein structure was checked for its structural accuracy by the following program: Ramachandran Plot Analysis, using SAVES SERVER (Structural analysis and verification server). Ramachandran plot Analysis was performed to determine the stability of the modelled structure. Subsequently the model structure was validated using PROCHECK, which determine stereo chemical aspects along with main chain and side chain parameters with comprehensive analysis.

Thrombomodulin (P07204) was subjected to homology search against PDB database using BlastP to identify significant structural homologues to be used as template for homology modelling. The results indicated the presence of CLECT super family domain and the best homolog was 3P5B with 42 % identity with the query protein and thus served as a template for modelling and the modelled protein obtained is shown in (fig.11.) and Validation was done using Ramachandran map (fig.12.) after loop refinement.

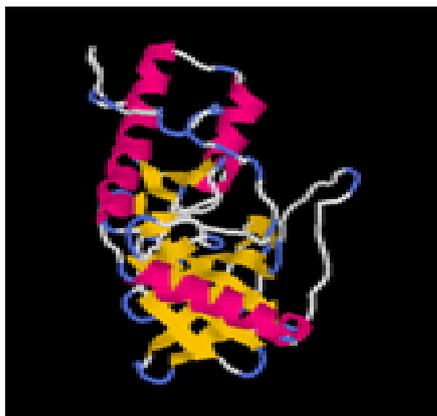


Fig. 9: Structure of Polo like kinase I visualised using Rasmol

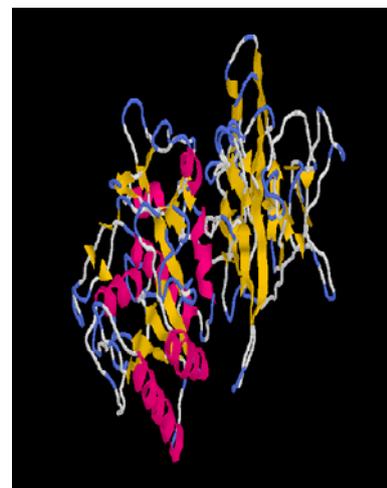


Fig. 11: Structure of Thrombomodulin visualised using Rasmol

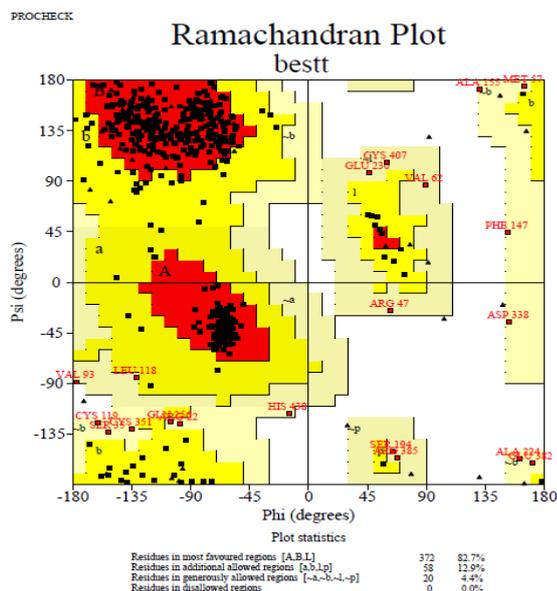


Fig. 12: Ramachandran plot of Thrombomodulin

Gene structure prediction

Over the past few years, there has been a gradual increase in both the accuracy of comparative models and the fraction of protein sequences that can be modeled with useful accuracy. The magnitude of errors in fold assignment, alignment, and the modeling of side chains, loops, distortions, and rigid body shifts has decreased measurably. This is a consequence of both better techniques and a larger number of known protein sequences and structures. Nevertheless, all the errors remain significant and demand future methodological improvements. In addition, there is a great need for more accurate detection of errors in a given protein structure model. Error detection is useful both for refinement and interpretation of the models. The biological role of a protein is determined by its function, which is in turn largely determined by its structure. Thus there is more benefit in knowing the three dimensional structures of all the proteins. Although more and more structures are determined experimentally at an accelerated rate, it is simply not possible to determine all the protein structures from experiments. As more and more protein sequences are determined, there is more need for predicting protein structures computationally. Decades of intense research in this area brought about huge progress in our ability to predict protein structures from sequences only. This process is an efficient way for enriching potential target genes, and for identifying those that are critical for normal cell functions.

Protein structure prediction aims to model the three-dimensional (3D) structure of so far structurally uncharacterized proteins from their amino acid sequence. Motivated by the observation that homologous proteins with related amino acid sequences have similar 3D structures, protein homology modelling uses comparative methods to generate models for a target protein based on one or more related proteins with known 3D structure. The coordinates of the model are generated based on alignments between the target's and template's amino acid sequences, which define the correspondence between residues in both proteins. Ultimately, the quality of a computational model determines its usefulness for specific biomedical applications. Therefore, model quality estimation methods are used to identify unreliable or erroneous regions in the resulting models, and to estimate the overall accuracy of a model.

Genscan

GENSCAN is a gene structure prediction system program which analyzes genomic DNA sequences from a variety of organisms including human, other vertebrates, invertebrates and plants. For each sequence, the program determines the most likely "parse" (gene structure) under a probabilistic model of the gene structural

and compositional properties of the genomic DNA for the given organism. This set of exons/genes is then printed to an output file (the text output) together with the corresponding predicted peptide sequences. A graphical (PostScript) output may also be created which displays the location and DNA strand of each predicted exon. Unlike the majority of other currently available gene prediction programs, the model treats the most general case in which the sequence may contain no genes, one gene, or multiple genes on either or both DNA strands and partial genes as well as complete genes are considered. The most important restrictions are that only protein coding genes are considered (and not tRNA or rRNA genes, for example), and that transcription units are assumed to be non-overlapping. [91]. Following are the GENSCAN results predicted for the proteins Plk-1 and Thrombomodulin.

Homo sapiens polo-like kinase 1

> Homo sapiens polo-like kinase 1 (PLK1), mRNA

> Homo sapiens polo-like kinase 1 (PLK1), mRNA

GAGCGGTGCGGAGGCTCTGCTCGGATCGAGGTCTGCACGGCAGCTTCGG
GAGCATGAGTGTGCAGTGAC

TGCAGGAAGCTGGCACGGGACCGGCCGACCTGGGAAAGCCGGGGTC
CCCGAGTTGCAGTCCCGGA

GCTCCGGCGGGCTCCACCGCGAAAGAGATCCCGGAGGTCTAGTGG
ACCCAGCAGCCGGCGCGCT

ATGTGCGGGCCGCTTTTGGGCAAGGGCGGCTTTGCCAAGTGCTTCGA
GATCTCGGACCGGACACCAA

GGAGTGTTCGCGGGCAAGATTGTGCCTAAGTCTCTGCTGCTCAAGCCG
CACCAGAGGAGAAGATGTCC

ATGGAATATCCATTACCGCAGCCTCGCCACCAGCAGTCTGATAGGAT
TCCACGGCTTTTTCGAGGACA

ACGACTTCGTGTTCTGTGGTGTGGAGTCTGCCGCGGAGGTCTCTCT
GGAGTGCACAAGAGGAGGAA

AGCCCTGACTGAGCCTGAGGCCGATACTACCTACGGCAAATTGTGCTT
GGCTGCCAGTACCTGCACCGA

AACCAGTATTTCATCGAGACCTCAAGCTGGCAACCTTTTCTGTAATG
AAGATCTGGAGTGAAAATAG

GGGATTTGGACTGGCAACCAAGTCAATATGACGGGGAGAGGAAGA
AGACCCTGTGTGGACTCTCTAA

TTACATAGTCCCGAGGTGCTGAGCAAGAAGGGCACAGTTTCGAGGT
GGATGTGTGGTCCATTGGGTGT

ATCATGTATACCTTGTGTAGTGGGCAAACACCTTTTGTGACTTCTTGCC
TAAAAGAGACCTACCTCCGGA

TCAAGAAGAATGAATACAGTATTCCCAAGCACATCAACCCCTGGCCGC
CTCCCTCATCCAGAAGATGCT

TCAGACAGATCCACTGCCCCCAACCATTAACGAGCTGCTTAATGAC
GAGTCTTTACTTCTGGCTAT

ATCCCTGCCCGTCTCCCATCAGCTGCCTGACCATTCCACCAAGGTTTTC
GATTGCTCCAGCAGCCTGG

ACCCAGCAACCGGAAGCCCTCACAGTCTCAATAAAGGCTTGAGAGAA
CCCCCTGCCTGAGCGTCCCCG

GGAAAAAGAAGAACCAGTGGTTCGAGAGACAGGTGAGGTGGTGGACTG
CCACCTCAGTGACATGCTGCAG

CAGCTGCACAGTGTCAATGCCTCAAGCCCTCGGAGCGTGGGCTGGTCA
GGCAAGAGGAGGCTGAGGATC

CTGCCTGCATCCCATCTTCTGGGTGAGCAAGTGGGTGGACTATTGGA
CAAGTACGGCCTTGGGTATCA

GCTCTGTGATAACAGCGTGGGGGTGCTCTTCAATGACTCAACAGCCTC
ATCCTTACAATGATGGTGAC

AGCCTGCAGTACATAGAGCGTGACGGCACTGAGTCTACCTACCGTGA
GTTCCATCCAACCTCTTGA

TGAAGAAGATCACCTCCTTAAATATTTCCGCAATTACATGAGCGAGC
 ACTTGCTGAAGGCAGGTGCCAA
 CATCACGCCGCGGAAGGTGATGAGCTCGCCCGGCTGCCCTACCTACGG
 ACCTGGTTCCGCACCCGACAG
 GCCATCATCTGCACCTCAGCAACGGCAGCGTGCAGATCAACTTCTTCC
 AGGATCACACCAAGCTCATCT
 TGTGCCACTGATGGCAGCCGTGACCTACATCGACGAGAAGCGGGACTT
 CCGCACATACCGCTGAGTCT
 CCTGGAGGAGTACGGCTGCTGCAAGGAGCTGGCCAGCCGGCTCCGCTAC
 GCCGCACTATGGTGGACAAG
 CTGCTGAGCTCACGCTCGGCCAGCAACCGTCTCAAGGCCTCCTAATAGC
 TGCCCTCCCTCCGGACTGGT
 GCCCTCCTCACTCCCACCTGCATCTGGGGCCATACTGGTTGGCTCCCGC
 GGTGCCATGTCTGCAGTGTG
 CCCCCAGCCCGGTGGCTGGGCAGAGCTGCATCATCTTGCAGGTGGG
 GGTGCTGTGTAAGTTATTTT
 TGTACATGTTGGGTTGGGTTTACAGCCTTGTCCCTCCCTCAA
 CCCACCATATGAATTGTACA
 GAATATTTCTATTGAATTCGGAAGTGTCTTTCCTTGGCTTTATGCACA
 TTAACAGATGTGAATATTTCA
 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

Gene: Homo sapiens polo-like kinase 1 (PLK1), mRNA

Tool: Genscan 1.0

Genscan Result

Predicted genes/exons

Gn. Ex Type S. Begin. End. Len Fr Ph I/Ac Do/T CodRg P. Tscr.

 1.01 Sngl + 93 1904 1812 2 0 76 38 2663 0.999 254.58

Predicted peptide sequence(s)

>GENSCAN_predicted_peptide_1|603_aa

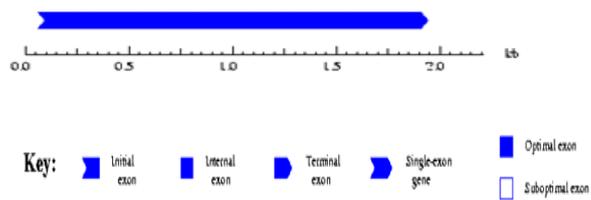
MSAAVTAGKLARAPADPGKAGVPGVAAPGAPAAAPPAKEIPEVLVDPRS
 RRRYVRGRFLG
 KGGFAKCFEISDADTKVEFAGKIVPKSLLKPHQREKMSMEISIHRS
 LAHQHVVGFGHGF
 EDNDFVFFVLELCRRRSLLELHKRRKALTEPEARYYLRQIVLGCQYLHR
 NRVIHRDLKLG
 NLFNLEDLEVKIGDFGLATKVEYDGERKKTLCGTPNYIAPEVLSKKGHSF
 EVDVWSIGCI
 MYTLLVGKPPFETSCLKETYLRKIKNEYSIPKHINVAASLIQKMLQTD
 P
 TARPTINELL
 NDEFFTSGYIPARLPITCLTIPPRFSIAPSSLDPSNRKPLTVLNKGLNPLP
 ERPREEKEE
 PVVRETGEVVDCHLSMMLQLHVSNAKPSERGLVRQEEAEDPACIPIF
 WVSKWVDYSDK
 YGLGYQLCDNSVGLFNDSTRLLIYNDGDSLQYIERDGTESYLVSSHPN
 SLMKKITLLK
 YFRNYMSEHLLKAGANITPREGDELARLPYLRTWFRTRSAILHLSNGSV
 QINFFQDHTK
 LILCPLMAAVTYIDEKRDFRTRYRLSLEEYGCCKELASRLRYARTMVDKL
 LSSRSASNRL KAS

Graphical Output

Inference

The Homo sapiens polo-like kinase 1 (PLK1), mRNA sequence have only the Single-exon gene (ATG to stop).

GENSCAN predicted genes in sequence Noname



Homo sapiens platelet factor 4- Thrombomodulin.

> Homo sapiens platelet factor 4 (PF4), mRNA

CCATCGCACTGAGCACTGAGATCCTGCTGGAAGCTCTGCCGAGCATGA
 GCTCCGAGCCGGTTCTGCG
 CCTCAGCCCCGGGTGCTGTTCTCTGGGGTTGCTGCTCCTGCCACTTGT
 GGTCGCCCTCGCCAGCGCTGA
 AGCTGAAGAAGATGGGACCTGCAGTGCCTGTGTGTGAAGACCACCTCC
 CAGGTCGGTCCACAGCACATC
 ACCAGCCTGGAGGTGATCAAGGCCGGACCCACTGCCCACTGCCAAC
 TGATGCCACGCTGAAGAATG
 GAAGGAAAATTTGCTTGGACCTGCAAGCCCCGCTGTACAAGAAAATAA
 TTAAGAACTTTTGGAGAGTTA
 GCTACTAGCTGCCTACGTGTGTGCATTTGCTATATAGCATACTTCTTTT
 TTCCAGTTTCAATCTAACTGT
 GAAAGAACTTCTGATATTTGTGTTATCCTTATGATTTAAATAAACAA
 AATAAATC

Gene: Homo sapiens platelet factor 4 (PF4), mRNA

Tool: Genscan 1.0

Result

Predicted genes/exons

Gn. Ex Type S. Begin. End. Len Fr Ph I/Ac Do/T CodRg P. Tscr.

 1.01 Sngl + 84 389 306 2 0 79 37 348 0.729 22.98

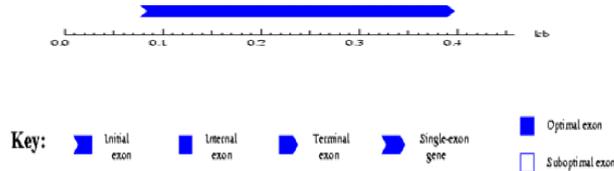
Predicted peptide sequence(s)

>GENSCAN_predicted_peptide_1|101_aa

MSSAAGFCASRPGLLFLGLLLLPLVAVAFASAEAEEDGLQCLCVKTTSSQV
 RPRHITSLEV
 IKAGPHCPTAQLIATLKNRGIKICLDLQAPLYKKIHKLLLES

Graphical output

GENSCAN predicted genes in sequence Noname

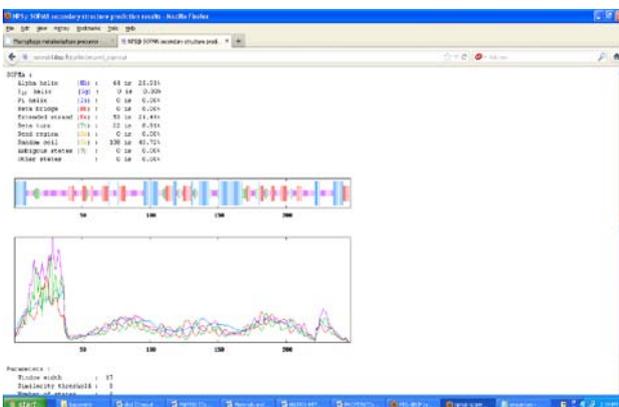
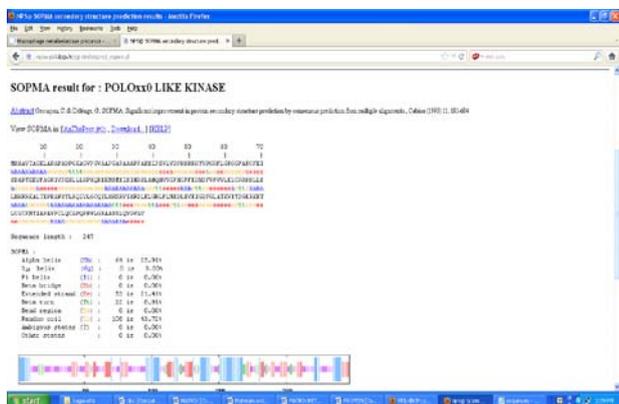


Inference

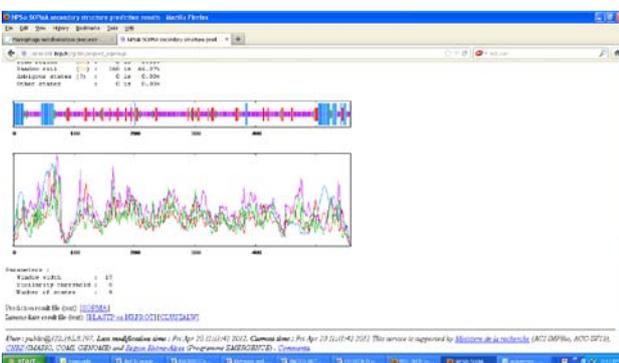
The Homo sapiens platelet factor 4 (PF4), mRNA sequence have Single-exon gene (ATG to stop), Poly-A signal (consensus: AATAAA).

SOPMA For secondary structure prediction

Self optimized prediction method is based on the homologue method of Levin *et al.*(1993). This method correctly predicts 69.5% of amino acids for a three description of the secondary structure (alpha-helix, beta-sheet and random coil) in a whole database containing 126 chains of non homologous proteins of SOPMA. Protein secondary structure prediction determines the regions of secondary structure in a protein at the level of alpha –helix, beta sheet and random coil, from information present in the primary protein sequence.



Parameters, Window width- 17, Similarity threshold-2, Number of states- 4



Parameters, Window width- 17, Similarity threshold-9, Number of states- 4

The predicted secondary structure of the PLK-1 in Lung cancer consisted of 25.91% predicted α-helices and β-sheets and 43.72% random coils and 8.91% turns were reported by Birve et. al. In

contrast to this it was significant to note that there was absence of β –bridges while α-helix, β-turns, coils and extended strands were present on different amino acid residues.

The predicted secondary structure of thrombomodulin in Lung cancer consisted of 14.54% predicted α-helices and β-sheets and 66.0% random coils and 6.10 turns % were reported by Birve et. al. In contrast to this, it was significant to note that there was absence of β –bridges while α-helix, β-turns, coils and extended strands were present on different amino acid residues.

Drug design and virtual ligand screening

Drugs are essential for the prevention and treatment of disease. Human life is constantly threatened by many diseases. Therefore, ideal drugs are always in great demand. To meet the challenges of ideal drugs, an efficient method of drug development is demanding. But the process of drug design, development and commercialization is a tedious, time-consuming and cost-intensive process. To fulfill these challenges, several multidisciplinary approaches are required for the process of drug development; collectively these approaches would form the basis of *In Silico* approach in drug design. A drug target is a biomolecule which is involved in signaling or metabolic pathways that are specific to a disease process. Biomolecules play critical roles in disease progression by communicating through either protein-protein interactions or protein-nucleic acid interactions leading to the propagation of signaling events and/or alterations of metabolic processes. Therefore, modulation of biological functions performed by these bio molecules would be potentially beneficial and could be achieved either

- i) by inhibiting their function with small molecules whose competitive binding affinity would be greater than their natural ligands that bind to the active sites or
- ii) by inhibiting the bio molecular interactions (between the biomolecules) by small molecules or
- iii) by activating bio molecules that are functionally deregulated in some disease such as cancer

Developing a lead molecule and an effective drug is challenging even for known targets. Recently, drug discovery has significantly increased due to the availability of 3D X-ray or NMR structures of biomolecules, docking tools, and the development of computer aided methodologies. Considering both the potential benefits to human health and the enormous costs in time and money of drug discovery, any tool or technique that increases the efficiency of any stage of the drug discovery enterprise will be highly prized.

In Silico drug designing is one of these tools which can be used to increase the efficiency of the drug discovery process. This approach cannot maximize its utility alone; rather it can form a valuable partnership with the experimentalist. It provides valuable information and helps to guide further experimental planning and potentially makes this process more efficient. *In Silico* Drug Designing process comprises of 3 major stages (Fig 14)

Stage 1: It involves identification of therapeutic target and building a heterogeneous small molecule library to be tested against it. This is followed by the development of avirtual screening protocol initialized by docking of small molecules from the library.

Stage 2: These selected hits are checked for specificity by docking at binding sites of other known drug targets.

Stage 3: These selected hits are subjected to detail *InSilico* ADMET profiling studies and those molecules that pass these studies are termed as leads.

Target identification

Target identification is the first key stage in the drug discovery pipeline. However, identification of drug able targets form among thousand of candidate macromolecules is still a challenging task. This can be achieved by extensive literature referring, pathway analysis and also by genomic and proteomics approaches for examples by comparison of the protein expression profiles [92].

Target validation

After a drug target has been identified, a rigorous evaluation needs to occur to demonstrate that modulation of the target will have the desired therapeutic effect.

Target validation process includes determining if the modulation of a target's function will yield a desired clinical outcome. *In Silico* characterization can be carried by using approaches such as genetic-network mapping, protein-pathway mapping, and protein-protein interactions

Lead discovery

The identification of small molecule modulators of protein function and the process of transforming these into high-content lead series are key activities in modern drug discovery.

Lead can be identified by one or more of several technology-based approaches like structure based design, virtual High-Throughput screening, literature and patent-based innovations [93].

Lead optimization

Lead optimization is the complex, no-linear process of refining the chemical structure of confirmed Lead molecules to improve its drug characteristics with the goal of producing drug candidate. Lead structures are optimized for target affinity and selectivity. Docking techniques are currently applied to aid this process [94].

Results of target structure prediction and active site analysis

The 3-D crystal structure of the targeted lung cancer protein Polo Like Kinase 1 and Thrombomodulin was modeled using suitable templates from the protein data bank (PDB) (www.rcsb.org/pdb). Structural and active site studies of the protein were done by using Q site finder and visualized using Rasmol software.

Ligand preparation and optimization

Using Chemsketch Software the structures of the drugs and analogs were sketched and generated their MOL File followed by subsequent generation of their 3-D structures by using molecule format converter tool. (fig.13). Inhibitors drawn in ACD/ChemSketch (Freeware).

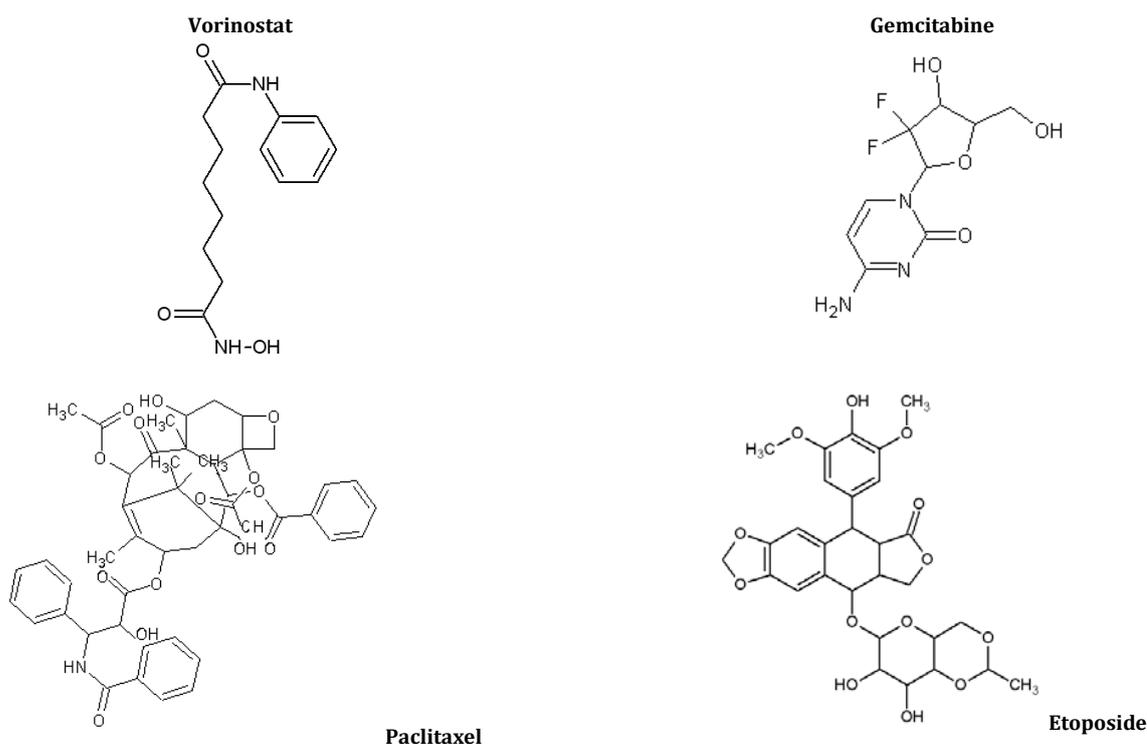


Fig.13: Inhibitors drawn in ACD/Chem Sketch (Freeware)

Receptor grid generation

Receptor grids were calculated for polo like kinase 1 and Thrombomodulin such that various ligand poses bind within the predicted active site during docking.

Active site prediction

After obtaining the final model, the possible binding sites of Polo like kinase 1 and Thrombomodulin were searched using Q-SiteFinder. Binding sites (<http://bmbpcu36.leeds.ac.uk/qsitefinder/>) were obtained for Polo like kinase 1 from Q-SiteFinder.

These binding sites were compared to the active site of the template to determine the residues forming the binding pocket as shown in the fig.14 & 15.

Active sites

LEU59, GLY60, LYS61, CYS67, ALA80, LYS82, GLU101, HIS105, VAL114, LEU130, GLU131, LEU132, CYS133,

ARG134, ARG135, ARG136, SER137, GLU140, GLY180, ASN181, PHE183, GLY193, ASP194, PHE195, VAL161,



3D structure visualized using rasmol

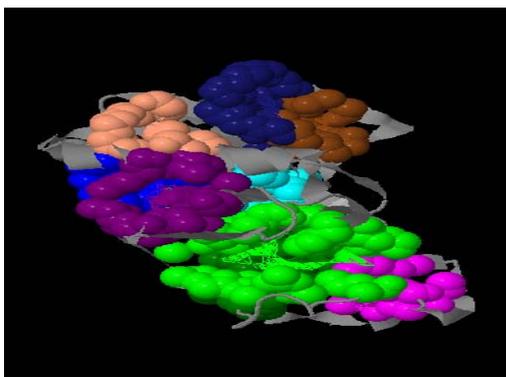


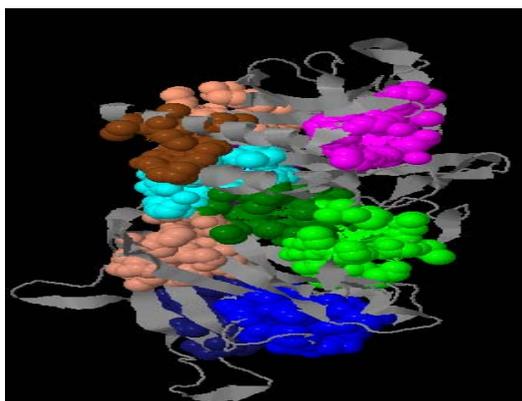
Fig. 14: Active site prediction using Q-sitefinder

LEU162,CYS164,GLN165,LEU167,HIS168,VAL172,ILE173,HIS174,ARG175,ASP176,LEU177,PRO215,

TYR217, ILE218, ALA219, PRO220, ALA221, PRO223, ARG232,LEU234



Active site prediction using q-sitefinder



Active sites

GLU141,ASP145,GLY146,PHE147,LEU148,CYS149,GLU150,PHE151,VAL171,SER172,ILE173,ILE241,

PRO245, GLN248, LEU255, GLN256, ALA257, GLY259, ARG260,

THR263, TYR296, VAL323, ASN324, THR325

Fig. 15: 3D model visualized using rasmol

Docking using autodock

The molecular docking was performed using Auto Dock; a suite of automated docking tools. The software is used for modeling flexible small molecule such as drug molecule binding to receptor proteins of known three dimensional structures. It uses Genetic Algorithms

for the conformational search and is a suitable method for the docking studies. The technique combines simulated annealing for conformation searching with a rapid grid based method of energy evaluation. Auto Dock tools is used to prepare, run and analyze the docking simulations, in addition to modeling studies. Auto Dock is the most cited docking software because it is very fast, it provides high quality predictions of ligand conformations and good correlations between inhibition constants and experimental ones. During the docking simulations, the inhibitors were regarded as flexible and subjected to an energy minimization. The ligand orientations were scored through the use of a force-field-based energy scoring function, and the top-scored binding structure was selected.

The vast literature review suggests specific drug targets against these proteins as Vorinostat, Gemcitabine, Paclitaxel and Etoposide. PUBCHEM reveals the biophysical properties of these compounds. Vorinostat has a Molecular Weight of 264.3202 [g/mol] with Molecular Formula $C_{14}H_{20}N_2O_3$, XLogP3: 1.9, H-Bond Donor: 3, H-Bond Acceptor: 3 and has SMILES notation as C1=CC=C(C=C1)NC(=O)CCCCC(=O)NO. Literature suggests the application of vorinostat in treatment of advanced non-small-cell lung cancer (NSCLC) that showed improved response rates and increased median progression free survival and overall survival. Gemcitabine has a Molecular Weight of 263.198146 [g/mol] with Molecular Formula $C_9H_{11}F_2N_3O_4$, XLogP3: -1.5, H-Bond Donor: 3, H-Bond Acceptor: 6 and has SMILES notation as C1=CN(C(=O)N=C1N)C2C(C(C(O2)CO)O)(F)F. Combination of gemcitabine and carboplatin has been found to be effective in treating several different types of cancer, but most commonly used to treat lung cancer. Paclitaxel has a Molecular Weight of 853.90614 [g/mol] with Molecular Formula: $C_{47}H_{51}NO_{14}$, XLogP3: 2.5, H-Bond Donor: 4, H-Bond Acceptor: 14 and SMILES notation as CC1=C2C(C(=O)C3(C(C4C(C3C(C2(C)C)C(C1OC(=O)C(C5=CC=C=C5)NC(=O)C6=CC=CC=C6)O)OC(=O)C7=CC=CC=C7)(CO4)OC(=O)C)OC(=O)C. Paclitaxel is approved in the UK for ovarian, breast and lung cancers and Kaposi's sarcoma. It is recommended in NICE guidance of June 2001 that it should be used for non small cell lung cancer in patients unsuitable for curative treatment, and in first-line and second-line treatment of ovarian cancer. Etoposide has Molecular Weight of 588.55658 [g/mol] with Molecular Formula $C_{29}H_{32}O_{13}$, XLogP3: 0.6, H-Bond Donor: 3, H-Bond Acceptor: 13 and has SMILES notation as CC1OCC2C(O1)C(C(C(O2)OC3C4OC(=O)C4C(C5=CC6=C(C=C35)OC6)C7=CC(=C(C(=C7)OC)O)OC)O)O. Etoposide phosphate is an anticancer agent. It is known in the laboratory as a topoisomerase inhibitor. It exploits the normal mechanism of action of the enzyme topoisomerase II, which aids in DNA unwinding and by doing so causes DNA strands to break. Cancer cells rely on this enzyme more than healthy cells, since they divide more rapidly. It is used as a form of chemotherapy for cancers such as Ewing's sarcoma, lung cancer, testicular cancer, lymphoma, non lymphocytic leukemia, and glioblastoma multiforme. It is often given in combination with other drugs. SMILES notation was drawn using ACD ChemsSketch and converted in to three dimensional PDB format using Molecular converter tool.

The 3D structures of these Lung Cancer proteins were docked with various inhibitors using Autodock Software. From the docking studies, it has been identified that Polo Like Kinase 1, Thrombomodulin, has been inhibited well by four drug compounds of the study.(FIG.16& 17.)

The Key interacting sites of Polo like Kinase 1 are LYS61, HIS168, HIS174, ARG175, LEU177, LYS178, GLY180, ASN181, ASP194, TYR217, ALA219. The active site is docked with the four drug compounds. Polo Like Kinase 1 interacts with Gemcitabine forming 6 Hydrogen bonds and binds strongly with a docking score of -7.98 Kcal/mol, Paclitaxel formed 5 Hydrogen bonds with a docking score of -6.86 Kcal/mol, Vorinostat forming 1 Hydrogen bonds and docking score of -8.43 Kcal/mol, with Etoposide forming 4 Hydrogen bonds and docking score of -8.57 Kcal/mol. Therefore, it can be seen that Gemcitabine is the most effective inhibitor of Polo like kinase 1. The Key interacting sites of Thrombomodulin are ALA168, VAL171, GLN256, ARG260, THR296. Thrombomodulin interacts with Gemcitabine forming 3 Hydrogen bonds and docking

score of -8.33 Kcal/mol, Paclitaxel forming 1 Hydrogen bonds and docking score of -7.69 Kcal/mol, Vorinostat forming 2 Hydrogen bonds and docking score of -10.1 Kcal/mol, Etoposide forming 2

Hydrogen bonds and docking score of -6.29 Kcal/mol. It is evident that Vorinostat acts as better inhibitor against Thrombomodulin receptor.

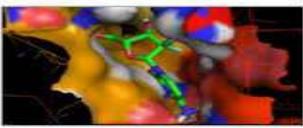
POLO LIKE KINASE 1				DOCKING SCORE (Kcal/mol)	H-BONDS	DOCKED STRUCTURE
LIGAND	RESIDUE	ATOM	ATOM			
GEMCITABINE	HIS174 LEU177 TYR217 ARG175 ARG175 ARG175	N N O O O	0 0 0 0 0	-7.99	5	
PACLITAXEL	GLY180 ASN181 LYS173 LYS81 LYS81	O N O O O	0 0 0 0 0	-8.88	5	
VORINOSTAT	ASP194	N	O	-8.43	1	
ETOPOSIDE	HIS188 ALA219 ARG175 ARG175	O N O O	0 0 0 0	-8.57	4	

Fig.16: Docking of polo like kinase 1 with potential inhibitors

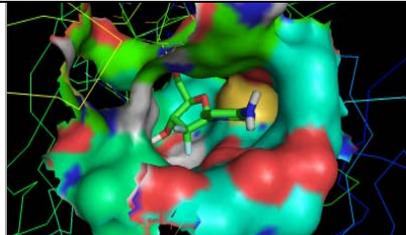
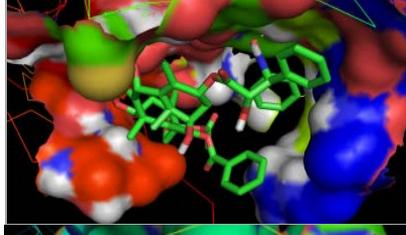
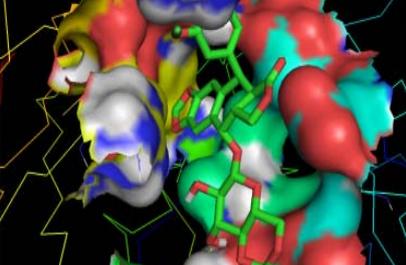
THROMBOMODULIN RESIDUE	ATOM	GEMCITABINE ATOM	DOCKING SCORE (Kcal/mol)	H-BONDS	
ARG260	N	O	-8.33	3	
GLN256	O	O			
VAL171	O	H			
THROMBOMODULIN RESIDUE	ATOM	PACLITAXEL ATOM	DOCKING SCORE (Kcal/mol)	H-BONDS	
THR294	OG1	O	-7.61	1	
THROMBOMODULIN RESIDUE	ATOM	VORINOSTAT ATOM	DOCKING SCORE (Kcal/mol)	H-BONDS	
GLN256 GLN256	O	O	-10.1	2	
THROMBOMODULIN RESIDUE	ATOM	ETOPOSIDE ATOM	DOCKING SCORE (Kcal/mol)	H-BONDS	
GLN203	O	O	-6.29	2	
ALA168	O	O			

Fig. 17: Docking of Thrombomodulin with potential inhibitors

These results suggest that all the four compounds are effective on their specific targets. The result of Lipinski's rule suggests that the drug targets are best therapeutic drugs. Docking study and *In silico* toxicity results proves the application of compounds as Potential and Natural therapeutic agents to treat Lung Cancer. Drugs based on this molecule could be useful against tumours with over expressed Plk1 and Thrombomodulin. Such drugs could selectively bind this kinase and thus lead to fewer side effects than a less selective drug. Molecules designed here formed stable bonds with PBD of Plk1 *in silico*. While laboratory syntheses of the molecules have not been done, it should be noted that experimentally known molecules were used as a model to design the protein. The merits of the designed molecules for anticancer applications eventually needs to be evaluated *in vitro* and, if warranted, *in vivo*. This study facilitates initiation of the drug discovery process for polo like kinase 1 and Thrombomodulin to present the scientific community with better inhibitors and/or drugs. Computational tools such as *in silico* docking provide the scientist with an alternate base for validation of lead molecules. In future, research work can be used further in clinical trials to test its effectiveness and for social benefits thus reducing the time and cost in drug discovery process.[95][96]

Perspectives

Many of the examples and applications discussed in this review indicate that the scoring and reliable ranking of test compounds continue to be major bottlenecks in structure-based virtual screening and lead optimization. Despite a plethora of already available scoring functions, further progress will be required to better account for and balance entropic effects and electrostatic interactions. Many current limitations are the result of the assumption that implemented solvation or entropic and electrostatic terms are generally applicable and transferable to different protein systems. However, structure based screening calculations have produced impressive results and many novel hits. These successes are at least in part due to the fact that virtual-screening campaigns mostly aim at the enrichment of active compounds, rather than, for example, accurate calculation of binding energies. For efficient compound selection, relying solely on computed scores is currently not sufficient; experience and intuition are often still a key to success. Taking this into account, further progress can be made in establishing more advanced scoring schemes, even if it is not possible to develop conceptually novel scoring functions in the near future. Importantly, scoring schemes can be advanced by modifying molecular systems used for benchmarking, calibrating selected functions for specific applications, or determining the most relevant scoring ranges. The statistical analysis of score distributions resulting from docking of large compound databases into different target sites has enabled scoring ranges to be determined that are most likely to reflect 'nonspecific' binding events.[97]

Concluding remarks

As also discussed in this article, although docking and scoring relies on many approximations, the application of these techniques during lead optimization, often in concert with other computational methods, already extends more traditional approaches to structure-based design. *In silico* cancer modeling presents significant opportunities to investigate oncogenesis across biological scales and systems. We are sure that these powerful methods will help to accelerate the development of diagnostic and therapeutic technologies for clinical medicine. Considerable improvement in the resolution, scale, and predictive power of these models must first be achieved, regarding which substantial challenges remain. Nonetheless, it may ultimately be possible to simulate oncogenesis and malignant invasion accurately from the scale of genetics to physiology. With this technology, distinguishing signatures of cancers could be discovered automatically for purposes of early diagnosis, prognosis, and treatment planning. Additionally, and perhaps most promisingly, with reliable digital representations of cancer, the effects of therapeutic interventions at both the molecular and surgical scales could be predicted *in silico* without exposing patients to risk. This innovation would greatly accelerate the development of safe and targeted anticancer therapeutics, and offer hope of medical treatments for diseases that remain refractory to

current clinical technologies. Ultimately, while confronting substantial experimental and analytical challenges, *in silico* models of cancer are advancing, and promise to strongly enhance both the fundamental understanding of cancer and its treatment in the clinic.

In future

Having said this, the bioinformatics techniques are becoming increasingly more effective, quicker, and simpler to use, and the databanks are growing in size and diversity. So these approaches, if used appropriately, will help to close the information gap between sequence and structure and complement *in vitro* approaches to investigate molecular structure and function. Additionally, many useful lessons can be learned from critical assessments of *bona fide* protein structure predictions in light of their structures being solved. Furthermore, the methods described in protein structure prediction projects can show, by example, how not to over interpret results obtained from bioinformatics methods. In this article, we hope we have demonstrated bioinformatics methods to predict protein structure by using a practical approach. A change in the working ethic between informatics and chemistry teams is required to make an *in silico*-driven process function successfully. This is surely a rate determining step in the transition in workflow.

Drug designing is a very complex, expensive and time consuming process. Bioinformatics provide a huge support to overcome the cost and time context in various ways. Bioinformatics provides wide range of drug-related databases and softwares, which can be used for various purposes, related to drug designing and development process. Though, bioinformatics is still in the initial phase and presently facing some hurdles, they show enough potential to help drug development process in near future

Summary

Because of the significant advantages outlined above, the use of *in silico* methods has grown significantly in popularity over the past couple of years. Specifically, most pharma companies have adopted some type of virtual screening capability to complement HTS methods and it is accepted that the predictions made by these techniques represent a fast method to enrich a biological screen. Some companies have already successfully adopted a far more radical discovery method that uses *in silico* methods to replace HTS. Drug discovery is often thought of (quite negatively) as a series of never ending bottlenecks. Currently, the largest attrition rates are associated with lead discovery and first trial in man stages. If the attrition rate could be dramatically reduced at the lead discovery stage that would represent a significant step forward and a more cost-effective discovery process would result. The signs so far are that *in silico* structure-based drug design can help deliver this vision.

The Protein-Ligand interaction plays a significant role in structural based drug designing. Our approach in Molecular Docking analysis resulted in the identification of potential drug targets. In the present work, we have taken the four key targets that play a crucial role in lung cancer and listed the drugs that were used against Lung Cancer to study its efficacy. From this study report we can conclude that some of the modified drugs are better than the commercial drugs available in the market. These drugs can be tested in wet lab and research and can be further validated for clinical trials. This study facilitates initiation of the drug discovery process for Lung cancer to present the scientific community with better inhibitors and /or drugs. In future, research work can be used further in clinical trials to test its effectiveness and for social benefits thus reducing the time and cost in the drug discovery process.

REFERENCES

1. Hanahan D, Weinberg RA. The hallmarks of cancer. *Cell* 2000;100(1):57-70.
2. Land H, Parada LF, Weinberg RA. Tumorigenic conversion of primary embryo fibroblasts requires at least two cooperating oncogenes. *Nature* 1983;304(5927):596-602.
3. Lloyd AC, Obermuller F, Staddon S, Barth CF, McMahon M, Land H. converge to regulate cyclin/cdk complexes. *Genes Dev* 1997;11(5):663-77.

4. Fanidi A, Harrington EA, Evan GI. Cooperative interaction between c-myc and bcl-2 protooncogenes. *Nature* 1992;359(6395):554-6.
5. Lowe SW, Cepero E, Evan G. Intrinsic tumour suppression. *Nature* 2004;432(7015):307-15.
6. McMurray HR, Sampson ER, Compitello G, Kinsey C, Newman L, Smith B, et al. Synergistic response to oncogenic mutations defines gene class critical to cancer phenotype. *Nature* 2008;453(7198):1112.
7. Chuang HY, Lee E, Liu YT, Lee D, Ideker T. Network-based classification of breast cancer metastasis. *Mol Syst Biol* 2007;3:140.
8. Liu ET, Lemberger T. Higher order structure in the cancer transcriptome and systems medicine. *Mol Syst Biol* 2007;3:94.
9. Auffray C. Protein subnetwork markers improve prediction of cancer outcome. *Mol Syst Biol* 2007;3:141.
10. Neely KE, Workman JL. The complexity of chromatin remodeling and its links to cancer. *Biochim Biophys Acta* 2002;1603(1):19-29.
11. Seligson DB, Horvath S, Shi T, Yu H, Tze S, Grunstein M, et al. Global histone modification patterns predict risk of prostate cancer recurrence. *Nature* 2005;435(7046):1262.
12. Esteller M. Cancer epigenomics: DNA methylomes and histone-modification maps. *Nature* 2007;8:286-98.
13. Jones PA. DNA methylation and cancer. *Oncogene* 2002;21:5358-60.
14. Esteller M, Fraga MF, Paz MF, Campo E, Colomer D, Novo FJ, et al. Cancer epigenetics and methylation. *Sci* 2002;297(5588):1807-8.
15. Laird PW. Cancer epigenetics. *Hum Mol Genet* 2005;14(90001):65-76.
16. Cummings MD, RL Desjarlais, AC Gibbs, V Mohan, EP Jaeger. Comparison of automated docking programs as virtual screening tools. *J Med Chem* 2005;48:962-76.
17. Schneidman-Duhovny D, Nussinov HJ. Wolfson predicting molecular interactions in silico: ii. protein-protein and protein-drug docking. *Curr Med Chem* 2004;11:91-107.
18. Sperandio O, MA Miteva, F Delfaud BO. Villoutreix receptor-based computational screening of compound databases: the main docking-scoring engines. *Curr Protein Pept Sci* 2006;7:369-93.
19. Zhou Z, AK Felts, RA Friesner, RM Levy. Comparative performance of several flexible docking programs and scoring functions: enrichment studies for a diverse set of pharmaceutically relevant targets. *J Chem Inf Model* 2007;47:1599-608.
20. Evers A, G Klebe. Successful virtual screening for a submicromolar antagonist of the neurokinin-1 receptor based on a ligand-supported homology model. *J Med Chem* 2004;47:5381-92.
21. Kairys V, MX Fernandes, MK Gilson. Screening Drug-Like compounds by docking to homology models: a systematic study. *J Chem Inf Model* 2006;46:365-79.
22. Mohan V, AC Gibbs, MD Cummings, EP Jaeger, RL Desjarlais. Docking: successes and challenges. *Curr Pharm Des* 2005;11:323-33.
23. Rockey WM, AH Elcock. Progress toward virtual screening for drug side effects. *Proteins* 2002;48:664-71.
24. Lyne PD. Structure-based virtual screening: an overview. *Drug Discov Today* 2002;7:1047-55.
25. Iskar M, Zeller G, Zhao XM, Van Noort V, Bork P. Drug discovery in the age of systems biology: the rise of computational approaches for data integration. *Curr Opin Biotechnol* 2012;23:609-16.
26. Whittaker P. What is the relevance of bioinformatics to pharmacology? *Trend Pharmacol Sci* 2003;24:434-9.
27. Ortega SS, Cara LC, Salvadori MK. In silico pharmacology for a multidisciplinary drug discovery process. *Drug Metabol Drug Interact* 2012;27:199-207.
28. Song CM, Lim SJ, Tong JC. Recent advances in computer aided drug design. *Brief Bioinform* 2009;10:579-91.
29. Speck-Planche A, Cordeiro MN, Guilarte-Montero L, Yera-Bueno R. Current computational approaches towards the rational design of new insecticidal agents. *Curr Comput Aided Drug Des* 2011;7(4):304-14.
30. Chen YP, Chen F. Identifying targets for drug discovery using bioinformatics. *Expert Opin Ther Targ* 2008;12:383-38.
31. Katara P, Grover A, Kuntal H, Sharma V. In silico prediction of drug targets in *Vibrio cholerae*. *Protoplasma* 2011;248:799-804.
32. Yamanishi Y, Kotera M, Kanehisa M, Goto S. Drug-target interaction prediction from chemical, genomic and pharmacological data in an integrated framework. *Bioinformatics* 2010;26:246-54.
33. Loh M, Soong R. Challenges and pitfalls in the introduction of pharmacogenetics for cancer. *Ann Acad Med Singap* 2011;40:369-74.
34. Bajorath J, Stenkamp R, Aruffo A. Knowledge-based model building of proteins: concepts and examples. *Protein Sci* 1994;2:1798-810.
35. Blundell TL, Sibanda BL, Sternberg MJE, Thornton JM. Knowledge-based prediction of protein structures and the design of novel molecules. *Nature* 1994;326:347-52.
36. Johnson MS, Srinivasan N, Sowdhamini R, Blundell TL. Knowledge-based protein modelling. *CRC Crit Rev Biochem Mol Biol* 1994;29:1-68.
37. Sali A. Modeling mutations and homologous proteins. *Curr Opin Biotech* 1995;6:437-51.
38. Sanchez R, Sali A. Advances in comparative protein-structure modeling. *Curr Opin Struct Biol* 1994;7:206-14.
39. Chothia C, Lesk AM. The relation between the divergence of sequence and structure in proteins. *EMBO J* 1986;5:823-26.
40. Koehl P, Levitt M. A brighter future for protein structure prediction. *Nature Struct Biol* 1999;6:108-11.
41. Lesk AM, Chothia C. How different amino acid sequences determine similar protein structures: the structure and evolutionary dynamics of the globins. *J Mol Biol* 1980;130:225-70.
42. Fischer D, Eisenberg D. Assigning folds to the proteins encoded by the genome of *Mycoplasma genitalium*. *Proc Natl Acad Sci USA* 1997;94:11929-34.
43. Huynen M, Doerks T, Eisenhaber F, Orengo C, Sunyaev S, Yuan Y, et al. Homology-based fold predictions for *Mycoplasma genitalium* proteins. *J Mol Biol* 1998;280:323-26.
44. Jones DT. Gen Threader: an efficient and reliable protein fold recognition method for genomic sequences. *J Mol Biol* 1999;287:797-815.
45. Rychlewski L, Zhang B, Godzik A. Fold and function predictions for *Mycoplasma genitalium* proteins. *Folding Design* 1998;3:229-38.
46. Sanchez R, Sali A. Large-scale protein structure modeling of the *Saccharomyces cerevisiae* genome. *Proc Natl Acad Sci USA* 1998;95:13597-602.
47. Bairoch A, Apweiler R. The swissprot protein sequence databank and its supplement TrEMBL in. *Nucleic Acids Res* 1999;27:49-54.
48. Abola EE, Bernstein FC, Bryant SH, Koetzle TF, Weng J, Allen FH, et al. Protein data bank. In *Crystallographic Databases—Information, Content, Software Systems, Scientific Applications*. Bonn/Cambridge/Chester. Data Commission Int. Union of Crystallography; 1987. p. 107-32.
49. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The protein data bank. *Nucleic Acids Res* 2000;28:235-42.
50. Zhang ZT. Relations of the numbers of protein sequences families and folds. *Protein Eng* 1997;10:757-61.
51. Holm L, Sander C. Mapping the protein universe. *Sci* 1996;273:595-602.
52. Brooijmans N, Kuntz ID. Molecular recognition and docking algorithms. *Annu Rev Biophys Biomol Struct* 2003;32:335-73.
53. Gohlke H, Klebe G. Approaches to the description and prediction of the binding affinity of small-molecule ligands to macromolecular receptors. *Angew Chem Int Ed* 2002;41:2644-76.
54. Halperin I, Ma B, Wolfson H, Nussinov R. Principles of docking: an overview of search algorithms and a guide to scoring functions. *Proteins* 2002;47:409-43.

55. Burnett RM, Taylor JS. Darwin: a program for docking flexible molecules. *Proteins* 2000;41:173-91.
56. Norel R, Lin SL, Wolfson H, Nussinov R. Shape complementarity at protein-protein interfaces. *Biopolymers* 1994;34:933-40.
57. Norel R, Petrey D, Wolfson H, Nussinov R. Examination of shape complementarity in docking of unbound proteins. *Proteins* 1999;35:403-19.
58. Connolly ML. Analytical molecular surface calculation. *J Appl Cryst* 1983;16:548-58.
59. Connolly M. Solvent-accessible surface of proteins and nucleic acids. *Sci* 1983;221:709-13.
60. Norel R, Wolfson H, Nussinov R. Small molecular recognition: solid angles surface representation and shape complementarity. *Comb Chem High Throughput Screen* 1999;2:177-91.
61. Goodford PJ. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J Med Chem* 1985;28:849-57.
62. Sayle RA, Milner-White EJ. Rasmol-Biomolecular graphics for all. *Trends Biochem Sci* 1995;20:374-6.
63. Sutcliffe MJ, Haneef I, Carney D, Blundell TL. Knowledge based modeling of homologous proteins, Part I: Three-dimensional frameworks derived from the simultaneous superposition of multiple structures. *Protein Eng* 1987;1:377-84.
64. Sutcliffe MJ, Hayes FR, Blundell TL. Knowledge based modeling of homologous proteins, Part II: Rules for the conformations of substituted side chains. *Protein Eng* 1987;1:385-92.
65. Sanchez R, Sali A. Comparative protein structure modeling. Introduction and practical examples with modeller. *Methods Mol Biol* 2000;143:97-129.
66. Friend G. Whatlf: a molecular modeling and drug design program. *J Mol Graph* 1990;8:52-6.
67. Guex N, Diemand A, Peitsch MC. Protein modeling for all. *Trends Biochem Sci* 1999;24:364-7.
68. Brocklehurst SM, Perham RN. Prediction of the three-dimensional structures of the biotinylated domain from yeast pyruvate carboxylase and of the lipoylated H-protein from the pea leaf glycine cleavage system: a new automated method for the prediction of protein tertiary structure. *Protein Sci* 1993;4:626-39.
69. Levi F. Cancer prevention: epidemiology and perspectives. *Eur J Cancer* 1999;35(14):1912-24.
70. A Lee J, Rodriguez D, Dosemeci M, Albanes D, Hoover R, Blair A. Leisure-time physical activity and lung cancer: a meta-analysis. *Cancer Causes Control* 2005;16(4):389-97.
71. Rodriguez V, Tardon A, Kogevinas M, Prieto S, Cueto A, Garcia M, et al. Lung cancer risk in iron and steel foundry workers: a nested case control study in Asturias, Spain. *Am J Ind Med* 2000;38(6):644-50.
72. Philip Bonomi. Matrix metalloproteinases and matrix metalloproteinase inhibitors in lung cancer. *Seminars in Oncology* 2009;29(1):78-86.
73. Bhagavathi S. Analysis of Lung Cancer Micro array data identifies new potential genes targets for Inhibitor design. Anil Prakash *Int J Adv Biotechnol Res* 2012;3(4):824-34.
74. Bhagavathi S. In silico modeling and validation of differential expressed proteins in Lung Cancer. Anil Prakash *Asian Pacific J Tropical Disease*; 2012. p. S524-9.
75. Takai N, Hamanaka R, Yoshimatsu J, Miyakawa I. Polo-like kinases (Plks) and cancer. *Oncogene* 2005;24(2):287-91.
76. Sanga S, Frieboes H, Zheng X, Gatenby R, Bearer E, Cristini V. Predictive oncology: multidisciplinary, multi-scale in-silicomodeling linking phenotype, morphology and growth. *Neuroimage* 2007;37:120-34.
77. Kim B, Cheng H, Grishin N, Hor A. Web server to infer homology between proteins using sequence and structural similarity. *Nucleic Acids Res* 2009;37:532-38.
78. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: Architecture and applications. *BMC Bioinformatics* 2009;10:421.
79. Friedrich A, Ripp R, Garnier N, Bettler E, Deléage G, Poch O, et al. Blast sampling for structural and functional analyses. *BMC Bioinformatics* 2007;8:8-62.
80. Greer J. Comparative model-building of the mammalian serine proteases. *J Mol Biol* 1981;153:1027-42.
81. Altschul SF, Madden TL, Schaffer AA, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 1997;25:3389-402.
82. Piedra D, Lois S, Cruz X. Preservation of protein clefts in comparative models. *BMC Struct Biol* 2008;8:2.
83. Bairoch A, Boeckmann B, Ferro S, Gasteiger E. Swiss-Prot Brief Bioinform 2004;5:39-55.
84. The Uni Prot Consortium The Universal Protein Resource (UniProt). *Nucleic Acids Res* 2007;36:190-5.
85. Boeckmann B, Blatter C, Famiglietti L, Hinz U, Lane L, Roehert B, et al. *Comptes Rendus Biologies* 2005;328:882-99.
86. Dowlathabad M, Anuraj N, Mukesh Y, Showmy S, Disha P. Comparative modeling of methylenetetrahydrofolate reductase (MTHFR) enzyme and its mutational assessment: in silico approach. *Int J Bioinformatics Res* 2010;2(1):05-09.
87. Laskowski RA, Watson JD, Thornton JM. Pro Func: a server for predicting protein function from 3D structure. *Nucleic Acids Res* 2005;33:89-93.
88. Laskowski RA, Rullmann JA, MacArthur MW, Kaptein R, Thornton JM. AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J Biomol NMR* 1996;8:477-86.
89. Wiederstein M, Sippl MJ. Pro SA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res* 2007;35:407-10.
90. Lu'thy R, Bowie JU, Eisenberg D. Assessment of protein models with three-dimensional profiles. *Nature* 1992;356:83-5.
91. Burge C, Karlin S. Prediction of complete gene structures in genomic DNA. *J Mol Biol* 1997;268:78-94.
92. Darryl L, Scott M. In Silico technologies in drug target identification and validation. Taylor Francis Group, LLC; 2006.
93. Bleicher KH, Bohm Hj, muller K, Alanine AI. Hit and Lead generation: beyond high-throughput screening. *Nat Rev Drug Discov* 2003;2(5):369-78.
94. Robert AG. Hit and lead identification: Integrated technology based approaches. *Drug Discovery Today* 2006;3(4):367-75.
95. Bhagavathi S, Anil Prakash. Molecular modeling and drug discovery of potential inhibitors for anti cancer target gene PLK-Polo like Kinase1. *Int J Pharm Bio Sci* 2014;5(1):(B)342-52.
96. Bhagavathi S, Anil Prakash. Molecular docking of Lung cancer proteins against specific drug targets. *World J Pharm Res* 3(3):4248-62.
97. H Godden JW, Stahura FL, Bajorath J. Statistical analysis of computational docking of large compound databases to distinct protein binding sites. *J Comput Chem* 1999;20:1634-43.