Comparative Study of Docking Tools for Virtual High Throughput Screening (vHTS) Accuracy

Dissertation submitted to

CENTER FOR COMPUTATIONAL BIOLOGY & BIOINFORMATICS

School of Information Technology,

Jawaharlal Nehru University

In partial fulfillment of the requirements

For the award of the degree of

Master of Technology

Submitted by

CHETNA KUMARI

Under Supervision of

PROF. INDIRA GHOSH

SCHOOL OF INFORMATION TECHNOLOGY

JAWAHARLAL NEHRU UNIVERSITY

NEW DELHI – 110067

July, 2009

Center for Computational Biology & Bioinformatics
School of Information Technology
Jawaharlal Nehru University
New Delhi-110067, India

CERTIFICATE

It is to certify that the research work carried and data presented in this dissertation entitled **"Comparative Study of Docking Tools for Virtual High Throughput Screening (vHTS) Accuracy"** has been carried out by **Chetna Kumari** under my supervision, in the School of Information Technology at the Center for Computational Biology & Bioinformatics (CCBB), Jawaharlal Nehru University (JNU), New Delhi. This work is original and has not been used so far fully or partially for submission for any degree or diploma in any other university.

Supervisor

Prof. Indira Ghosh

Center for Computational Bio. & Bioinformatics,

School of Information Technology,

Jawaharlal Nehru University, New Delhi

Dean

Prof. Indira Ghosh

School of Information Technology,

Jawaharlal Nehru University,

New Delhi

Chetna Kumari

Student

Chetna Kumari

ACKNOWLEDGEMENT

My deepest gratitude goes to my supervisor and Dean Prof. Indira Ghosh, for accepting me as her student. I am thankful to her for the valuable time spent with me providing guidelines with her constructive ideas, valuable criticism and faith in me which really interested me to complete the work enthusiastically. She has always supported me, never letting let my failures sink in. Needless to say, I would have never reached this stage without her support.

I thank all the faculty members Prof. Alok Bhattacharya, Prof. Ram Ramaswamy, Prof. Rahul Roy, Prof. N. Parimala, Dr. Pradipta Bandopadhyay, Dr. A Krishnamchari, Dr. Andrew M Lynn, Dr. Supratim Sengupta, Dr. Dev Priya Choudhury, Dr. Narinder Singh Sahni, Dr. N. Subba Rao, Dr. Lovkesh Vig for being supportive during M.TECH course work and whenever I needed.

I am thankful to my seniors Om Prakash, Prashant and Amit for their timely help during my Project work.

I thank all the non-teaching members of the school for their help and support.

Staying in JNU is an experience I will cherish for the nice company of friends. I have had great times with them.

I acknowledge DBT for the financial support.

Chetna Kumari

Table of Contents

CHAPTER 1.

INTRODUCTION

Though today no drug can be designed completely *in silico*, Virtual High-Throughput Screening (vHTS) methods or High-Throughput Screening (HTS) *in silico,* comprising both bio-as well as Chemo-informatics is an indispensable and fundamental technology of pharmaceutical research. According to a survey of nearly 50 companies and academic institutions by Boston Consulting Group (BCG) **(http://www.bcg.com)** in November, 2001 [Tollman P et al., 2001] and approximately same estimate of drug development cost by DeMasi J A et al., 2003, a single drug discovery and development costs an average of $880 million and 15 years of research -from start to finish -i.e. from disease target identification and validation, through the discovery and optimization of lead structures, to clinical tests and regulatory approval (Fig1). Roughly 75% of these costs were attributed to failures. *In silico* methods save an average of $130 million (divided almost equally between biological target validation, searches for and optimization of chemical lead structures) and 0.8 years per drug [Seifert et al., 2003]. To maximize the output of drug discovery processes experimental (HTS) and theoretical approaches (vHTS) has often been integrated by introducing various statistical and filtering methods [Bajorath J, 2002]. To further improve the productivity of processes new technologies are being combined to decrease the timescale at every stage, so as to reach success or failure at faster rate [Collins I, 2006].

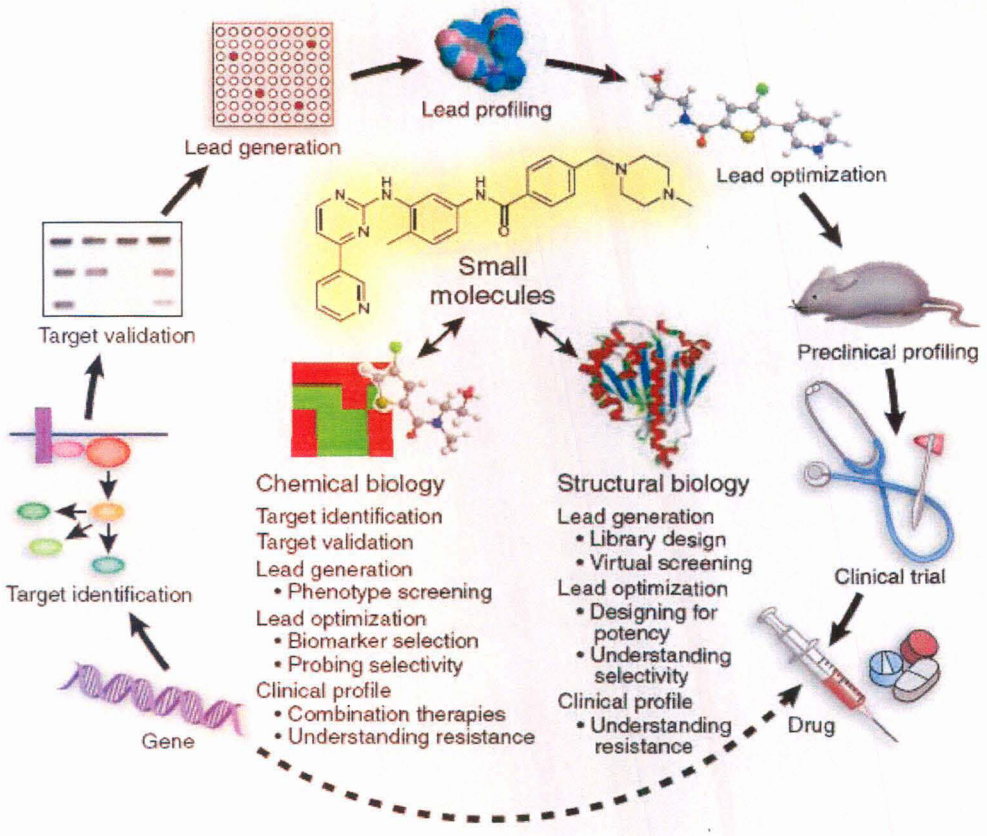


Fig1. Drug development cycle- from gene to drug [Collins I, 2006]

Incidence of breast cancer The age-standardized incidence rate of 37.4 per 100,000 makes breast cancer the most common malignancy among women worldwide (GLOBOCAN 2002, IARC) (http://www-dep.iarc.fr/globocon/globocon.html) (Fig 2). In India, the overall incidence rate is next to that of utero-cervical cancer (Fig 3) (GLOBOCAN 2002, IARC) while the consolidated report of population based cancer registries show breast cancer having the highest incidence rates in Delhi among all the cosmopolitan cities (ICMR, INDIA) (http://www-canceratlasindia.org/PriliminaryPages_1.html)

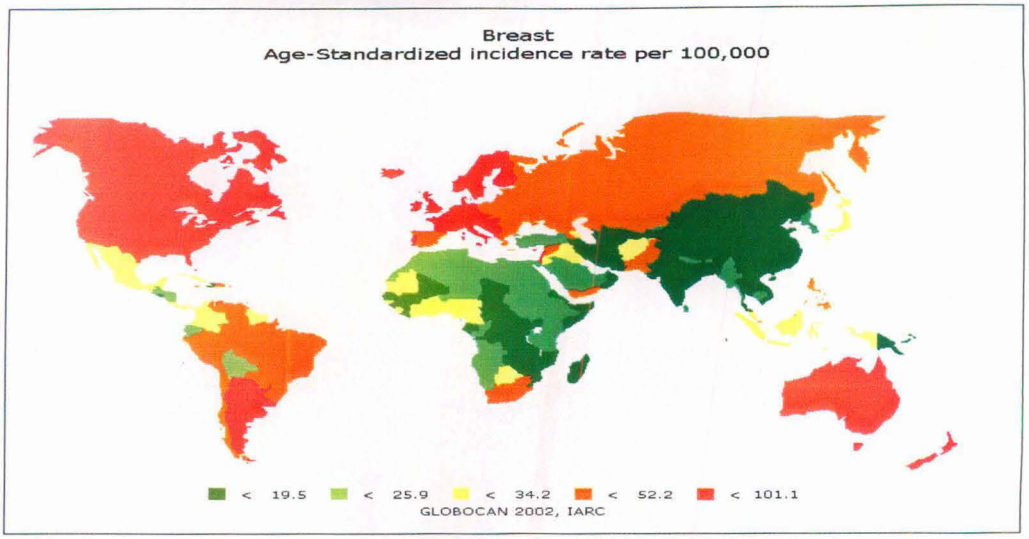


Fig 2. Age standardized incidence rate of breast cancer around the world.

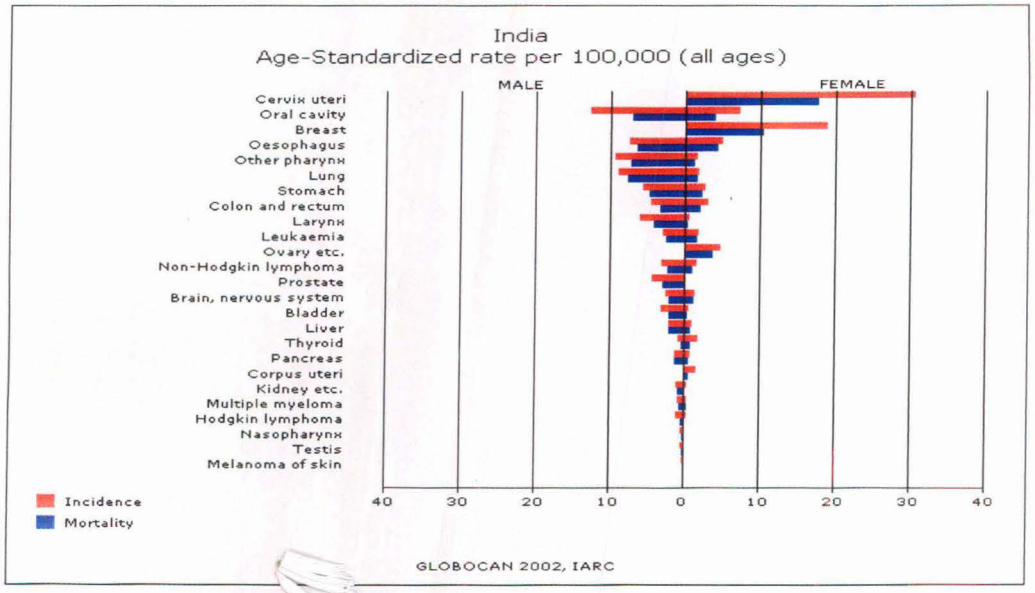


Fig 3. Age standardized incidence rate of different cancers in India

With this brief introduction in chapter 1, chapter 2 reviews on Virtual High-Throughput Screening (vHTS), Docking or Protein structure based compound screening, comparative study of docking tools; its past, present and future challenges which often proves to be bottleneck in the field of Computer-aided drug design (cadd).

Chapter 3 discusses Cancer (frequent cause of death in India and other part of world and 2nd most common cause of death in United State) and its therapeutic aspects.

Chapter 4 discusses case study BARD1- BRCT domain; a protein associated with BRCA1 tumor suppressor protein, dysfunctioning of which contributes to the development of Breast and ovarian cancer.

Chapter 5 is aims and objectives of the study i.e. to compare the performance of commonly used software for vHTS like DOCK, GOLD, GLIDE and FRED.

Chapter 6 discusses material & methods used for the study, highlighting search algorithms and scoring functions of the selected docking tools.

Chapter 7 discusses results of the work done and is divided into two parts: docking protocol and BARD1-BRCT domain.

Chapter 8 is discussion, conclusion and limitations of the study.

CHAPTER 2.

LITERATURE REVIEW

2.1 Virtual High Throughput screening (vHTS)

Computationally intensive in nature, Virtual Screening (VS) is one of the most popular among various theoretical approaches available today to complement the array of experimental high-throughput discovery technologies, sharing similar tasks and goals of searching large compound databases *in silico* and prioritizing a limited number of candidate molecules for testing to identify Novel Chemical Entities (NCE) having the desired biological activity [Bajorath J, 2002]. Hence, vHTS methods attempt to integrate computer science with biophysics using the synergy, the flexibility, cost-effectiveness, speed of computational algorithms and biophysical knowledge on molecular recognition.

Two main approaches of vHTS [Duhovny,2005] are Pharmacophoric approach [Joseph-McCarthy D, 2003] which is based on ligand structures and can be applied if the structure of the target receptor is not available while Docking methods or structure-based compound screening can be used if the receptor structure is known. Receptor-based VS faces several fundamental challenges [Moitessier N et al., 2008] i.e. sampling the various conformations of flexible molecules and calculating absolute binding energies in an aqueous environment. Even with its current limitations, VS accesses a large number of possible new ligand. Several success stories have been evidenced where prediction of a new ligand with their receptor-bound structures and in several cases with significantly greater hit rates (ligands discovered per molecules tested) than with experimental HTS [Sunderberg S.A, 2000, Stahura F.L & Bajorath J, 2004]. The major bottlenecks for biological screening are target validation and assay development. A list of targets addressed by VS is shown in table 1.

Table 1: Targets addressed by Virtual Screening [Kubiny H, 2006 & Klebes G, 2006]

G-protein-Coupled receptors	a1Aadrenoceptor,dopamineD3receptor,endothelinA (ETA) receptor, melanin-concentrating hormone type 1 receptor, muscarinic M3 receptor, neurokinin-1 receptor, neuropeptide Y receptor type5,purinergicA2Areceptor, urotensin II receptor (GPR14)
Nuclear receptors	Retinoic acid receptor, thyroid hormone receptor
Kinases	Akt 1 (also known as protein kinase Ba), Bcr-abl tyrosine kinase, checkpoint kinase 1, cyclin-dependent kinase 2, cyclin-dependent kinase 4, glycogen synthetase kinase, p56 lymphoid T cell tyrosine kinase, protein kinase CK2 (also known as casein kinase II),transforming growth factor b receptor kinase
Proteases	Cathepsin D, falcipain-2, HIV protease, plasmepsin II, severe acute respiratory syndrome CoV 3C-like proteinase, thrombin
Other hydrolases	Acetylcholinesterase, adenylate cyclase (oedema factor and CyaA, a toxin of the pathogenic bacteria Bacillus anthracis and Bacillus pertussis), AmpC b-lactamase, phosphodiesterase 4, protein tyrosine phosphatase 1B
Oxidases and reductases	Aldose reductase, dihydrofolate reductase, inosine 5` -monophosphate dehydrogenase inhibitors
Miscellaneous enzymes	5-Aminoimidazole-4-carboxamide ribonucleotide transformylase, carbonic anhydrase II, DNA gyrase,dTDP-6-deoxy-D-xylo-4-hexulose 3,5-epimerase, farnesyl transferase, guanine phosphoribosyl transferase, HIV-1 integrase, tRNA-guanine transglycosylase
Ion channels	T-type selective Ca2+ channel, Kv1.5 potassium channel, shaker potassium channel
Protein-2 interface & protein complexes	Bcl-2 protein–protein interaction, cyclophilin A, FK506-binding protein, mesangial cell proliferation,Rac1 protein–protein interaction, VLA-4 (also known as a4b1 antigen)
Protein-RNA interaction	HIV-1 RNA transactivation response element

2.2 Protein Structure Based Compound Screening or Docking

Docking section which was first introduced through CASP2 (Critical Assessment of Techniques for Protein Structure Prediction) experiment in 1994 [Dixon J.S, 1997], is a cornerstone technology of pharmaceutical industry for vHTS. Its two main ingredients are: search algorithm and scoring function (SF) with accuracy and algorithm execution time as its primary concerns. Docking tools are applied; A)**To screen the database of compounds** (of million-billion in number) **for a set of potential leads,** reducing the number of *in vitro* test B) **Inverse docking** [Chen Y et al., 2002] in which a single ligand is screened against numerous protein to find the differential binding ability. Potential applications of the later method in facilitating drug discovery include, (i) identification of unknown and secondary therapeutic targets of a drug, (ii) prediction of potential toxicity and side effect of an investigating drug and (iii) To check the specificity of the potential drug against homologous proteins and (iv) probing molecular mechanism of bioactive herbal compounds extracted from plants used in traditional medicines. C) **Predicting protein-protein, protein-DNA, protein-RNA** interactions D) to solve **protein folding** problems Halperin I et al., 2002].

Ongoing progress in crystallography and multidimensional NMR (Nuclear Magnetic Resonance) studies has benefitted Structure Based Drug Design (SBDD) by generating more number of 3D structure of many proteins especially enzymes. Docking small molecular weight ligand to therapeutically relevant macromolecule has become a major computational tool for predicting protein- ligand interaction and guiding lead optimization [Brooijmans N et al, 2003]. With the first success story of structure based design of Anti hypertensive drug Captopril, an Angiotensin-Converting Enzyme (ACE) inhibitor; the structure of which was designed rationally from binding site model, using 3D information of an inhibitor complex of the closely related zinc protease Carboxypeptidase A, the increase in number is evidenced by the anti-glaucoma agent Dorzolamide, the HIV protease inhibitors Nelfinavir and Amprenavir, and the neuraminidase inhibitor Zanamivir, others still being in preclinical or clinical development. Successful docking case which has popularized the field of SBDD is listed in table 2.

Table 2: List of some of the successful docking studies [Schulz-Gasch T, 2004].

Target	Tools	References
PTP1B(Protein Tyrosine Phosphate 1B)	Dock 3.5	Doman, T.N et al., 2002
HCA (Human Carbonic Anhydrase)	CombiSMoG	Gerzybowski,B.A et al., 2002
AmpC beta- lactamase	NW-Dock	Powar, R.A et al., 2002
tRNA-guanine transglycosylase	Flex-X 1.102	Brenk R, et al., 2003
IMPDH(InosineMonophosphate Dehydrogenase)	Flex-X	Pickett, S.D et al., 2003
CK2 (Casein Kinase II)	Dock 4.01	Vangrevelinghe, E et al., 2003
DHFR (Dihydrofolate Reductase)	Flex-X 1.7.0	Wyss, P.C et al., 2003
Chk-1 Kinase (Checkpoint-1 Kinase)	Flex-X-pharm	Lyne, P.D et al., 1968

2.3 Comparative Study of Docking Tools

From the pioneering work of Kuntz et al., 1982 various docking programs based on different physicochemical approximation have been reported [Taylor & Jewsberry et al., 2002, and Halperin I et al., 2002]. Some of the docking tools widely used in industries as well as academia are listed in table 3 [Schulz-Gasch T, 2004]. The recent literatures citing docking study is full of benchmark addressing three possible issues: i) the capability of algorithm to reproduce the **x-ray pose** of ligand ii) the propensity of fast scoring functions to predict binding free energies from the **best scored pose** [Baxter et al., 1998] iii) the discrimination of **known binders** from randomly chosen molecules in VS experiments [Zavodszky MI et al., 2002]. To date when over 60 docking programs and 30 scoring functions (SFs) have been disclosed and none of the docking programs is known to be completely outperform the others [Moitessier N et al., 2008], there is need to perform comparative studies of different docking tools. The criteria for selection of docking tools for comparative study depends on : i) Choice of protein active site ii) diverse algorithm tools for docking and iii) availability of chemical database for VS (database docking). However, the difficulties regarding comparative study lies in the availability of many diverse algorithm, independent studies assessing the relative performance of docking algorithms, and comparisons of the scoring functions using the same active site and different

procedures. Major publications focus on the use of many methods [Bissantz C et al., 2000] whereas the quality of analysis may vary depending on the examined properties like quality of -the top ranked pose, all possible poses, binding free energy prediction, and VS utility. The levels of approximation for most docking programs vary considerably [Halperin I et al., 2002] leading to very inhomogeneous docking time ranging from few seconds to few hours. Also many docking programs have been calibrated and validated on small protein-ligand data sets. Detailed benchmarks (> 100 PDB ligand complexes) are reported only for a few docking tools. [Diller et al., 2001, Verdonk M.L et al., 2003, Nissink et al., 2002, Kellenberger E et al., 2004]. The major drawback in the estimation of binding energy for ranking compounds is that experimentally measured Ki does not directly correlate with the empirical scoring functions, hence quantitative comparison between each of the docking methods using the same protein and database is still difficult. Table 3 shows the most popular programs with the information on their algorithm and specific scoring function etc. It will be worthwhile to compare the same protein with its known ligands to validate the docking protocol in the beginning using most of these methods and then use the same protocol for an unknown protein and compare the docking profiles of the best hit compounds. This effort will help to choose the suitable docking program for getting more true positive hits and minimize the true negatives while screening a compound library.

Table 3: Main characteristics of five key commercial docking tools* [Schulz-Gasch T, 2004]

Name	Dock	FlexX	FRED	Glide	Gold
Vendor	UCSF www.dock.compbio .ucsf.edu	BiosolveiT/Tripos www.biosolveit.de	OpeneyeScientific Software,www.eyesopen. com	SchrodingerInc ww.schroedinger.com	CCDC www.ccdc.cam. ac.uk
Docking Algorithm	Shapefitting (sphere sets)	Incremental construction	Shape fitting (Gaussian)	Core placement and internal generation of conformations (MonteCarlo sampling)	Genetic Algorithm
Scoring	Depending on version: -Contact (shape fitting) score - Chemscore -Force Field interaction energy (amber) -Electrostatic energy score (DelPhi) -GB/SA solvation scoring	-FlexX score -PLP -ScreenScore -Drugscore	- Gaussian shape score -ScreenScore - PLP - User defined scoring	-Glidescore (Chemscore-related, terms for electrostatic mismatch) - Glidecomp (adds Coulomb-vdW energy Score)	-Goldscore -Chemscore -User defined scoring
Characteristics	-Many versions available- Manysuccessful applications reported -Medium throughput	-Fast- Pharmacophore constraints Many evaluation studies& successful applications reported - Gets trapped for ligands with too many rotatable bonds - Problematic placement of base fragment for ligands with mainly hydrophobic interactions - Good results when key H-bond interaction are required	-Very fast - Consensus scoring - Multiple active site correction Pharmacophore constraints - Analysis tool for fast optimization of settings (FredA) - Requires good shape complementarities between ligand & receptor for good performance - Relatively new tool	-Medium throughput -Special input format required (*.mae) - Relatively new tool	-Partial protein flexibility -Many ex. for accurate binding mode prediction reported -Medium to low throughput
Key References	Kuntz et al., 1982	Kramer et al., 1996	McGann M.R et al., 2003	Halgren T.A et al., 2004	Jones et al., 1997

* Other commercial docking tools are Autodock [Morris et al., 1998], ICM [Trotov M et al., 1994], QSP/Flo+ [McMartin C et al., 1997], Surflex [Jain A.N, 2007].

2.4 Objectives of Comparison of Different Methods

Comparison of the efficiency of different docking software is required due to the rapid evolution of docking and scoring algorithms and their use in drug design. Comparative study can be classified into target-oriented studies [Ha Sookhee et al., 2000, Hu X et al., 2004]; aims to identify best program for a specific target and broad

comparative studies; aims to evaluate the programs' accuracy on a set of proteins. Broad comparative studies aim to evaluate and compare the ability of different docking programs and/ or SFs (either separately or in combination) to achieve three goals, (1) properly (re-dock the ligand like the crystal pose and evaluate using Root Mean Square Deviation or RMSD) dock compounds to proteins, was the yardstick of CASP II experiment. (2) predict the ligand binding affinities, (3) extract active compounds from libraries of decoys.

CASP

CASP (Critical Assessment of Techniques for Protein Structure-Prediction), a community-wide experiment taking place every two years since 1994. First CASP meeting was held in 1994 in Asilomar while the 8th meeting held on Dec. 3-7, 2008, in the Island of Sardinia, Italy (www.Predictioncenter.org/casp8). CASP provides an opportunity to assess the quality of methods for protein structure prediction from primary structure of the protein. These methods includes: (1) *ab initio* protein modelling, (2) Comparative protein modeling, (3) Side chain geometry prediction, (4) Software, (5) Protein-protein complexes. If target sequence is found (e.g. using sequence alignment method i.e. BLAST,FASTA) similar to a protein sequence of known structure, comparative modelling is used to predict 3D structure, otherwise protein threading or *de novo* protein structure prediction must be applied. In CASP experiments blind prediction of structure from amino acid sequence consists of three parts: 1.Collection of targets for prediction from experimental community: X-ray crystallographers and NMR Spectroscopists, 2. Collection of prediction from modelling community, 3. Assessment and discussion of the result.

Different phases of CASP experiments includes CASP I to CASP VIII out of which CASP II [Dixon J.S, 1997, Dunbrack et al.,1997] need to be focused as a fourth category, Docking section [Dixon et al.,1997] was introduced in addition to three categories:1. Comparative modelling, 2.Threading or fold recognition 3.*ab initio* prediction which was already included in CASPI experiment and a separate experiment CAPRI (Critical Assessment of Predicted Interactions) [Jannin J et al., 2003], which is a test of protein docking experiment like CASP, was set up.

Using docking program with different protocol

An analysis by Viet et al., 1998 of prevalent search techniques using the CHARMM (Chemistry at HARvard Macromolecular Mechanics) force field compared Molecular Dynamics (MD), MC (Monte Carlo) and a GA (Genetic Algorithm) in the docking of 5 complexes showed the performance of the different techniques depend on the size of the binding site. MD provided the best efficiency in docking structures in large space, whereas the GA is best for a small search space which seem to contradict other results; for example search algorithm in AutoDock 3.5 [Goodsell D S and Olsson AJ, 1990] has transferred from MC/simulated annealing to a GA to efficiently sample large search space.

Using known actives from libraries of decoys

Dock [Kuntz et al., 1982, Schiochet B.K et al., 1992, Meng E.C et al., 1993, Ewing T.J.A et al., 1997, 2001], FlexX [Rarey M et al., 1996, 1999] and GOLD [Jones G et al., 1997] were evaluated in combination with seven different sfs and two protein targets i.e. thymidine kinase (TK) and estrogen receptor (ER) along with two random databases of 990 ligands with 10 true hits in each [Bissantz C et al., 2000]. Out of all docking programs and sfs combinations GOLD gave the best RMSD solutions and best ranking of ligands for both TK and ER as reviewed by Taylor R.D & Jewsberry P et al., in 2002.

Proteins active sites and docking preferences

Structure-based virtual screening experiments for seven different targets that differ significantly in the characteristics of their binding sites roughly grouped the binding sites into three different classes: lipophilic buried cavities (COX-2, estrogen receptor), targets of intermediate polarity with hydrogen bonding motifs common to the majority of inhibitors (p38 MAP kinase, gyrase B, thrombin) and targets with very polar, solvent-exposed binding sites (neuraminidase, gelatinase A). The calculations were performed with a variety of different objectives and SFs in combination with three fast, docking programs FlexX, FRED and Glide. The results show clearly that the performance of docking algorithms, objective functions and SFs strongly depends on characteristics of the target structure [Schultz-Gasch T & Stahl M in 2003].

2.5 Limitations of Comparative Studies

Comparative studies provide insight into the various programs' accuracy, speed, applicability to a range of targets and other factors with several limitations: First, the versions of the programs in question which may vary from one study to another, are not always specified. Second, the accuracy is greatly dependent on the settings employed and fine-tuning parameters may lead to significant changes in accuracy. Third, the preparation of receptors and ligands for docking requires knowledge of active site and careful consideration of all possible isomers and protonation states for which expertise on both chemistry and programming is required. Fourth, using RMSD as a criterion of docking accuracy is questionable [Moitessier N et al., 2008]. For example, an RMSD of less than 2Å used as a criterion of success gives misleading results when specific interactions such as directional H-bond or solvent exposed groups are to be considered. Sixth, sometimes no/ insufficient mention is made regarding the CPU (Central Processing Unit) time required to run the program. As many programs (e.g. GOLD and Glide) provide various levels of speed and accuracy, a comparative study should reveal the accuracy to expect within a specified period of time. Overall, the best indicator of programs accuracy is its ability to identify novel compounds amongst the best ranked by vHTS and experimentally confirmed.

In spite of several efforts to the development of accurate and fast docking / scoring method a universal program is yet to be developed. Several issues that remain to be addressed are, treatment of protein flexibility and in particular the scoring of modelled protein conformations, presence of water molecules and inclusion of the evaluation of the binding free energy. Ligand forming covalent complexes and macromolecules such as nucleic acid and metal-containing enzymes are not studied enough using docking methods.

CHAPTER 3.

ROLE OF PROTEINS IN CANCER

Cancer generation is a multistep process in which any cell developing into malignant one become self sufficient in growth signals, insensitive to anti-growth signals, evade apoptosis, sustain angiogenesis, gain limitless replicative potential and acquire potential for tissue invasion and subsequent metastasis. This malignant process produces oncogenes with dominant gain-of-function and tumor suppressor genes with recessive loss-of-function. Except for some inherited genetic mutations cancer is caused mainly by mutation in somatic cells. Hence, Cancer is a genetic disease evolving from a single cell, involving dynamic changes in the genome in form of mutations in critical genes controlling cell divisions and cell death [Hanahan & Weinberg, 2000] (Fig 4).

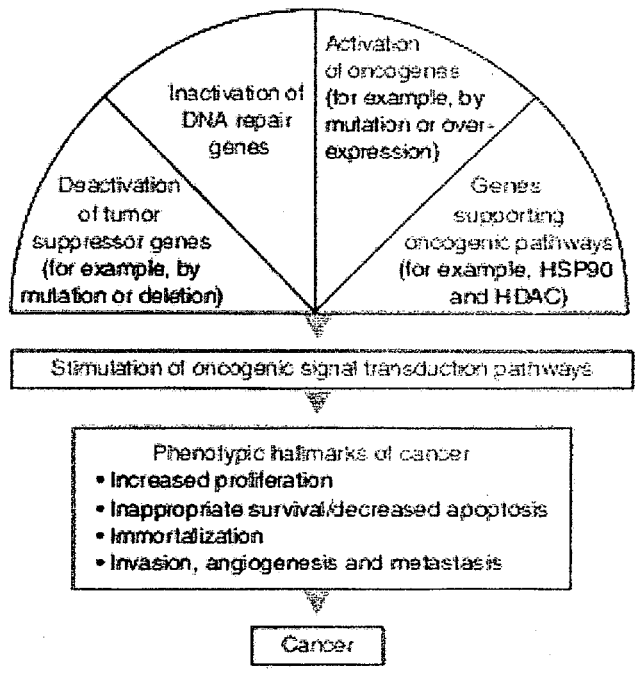


Fig 4: Hallmark of Cancer [Collins I, 2006]

3.1 DNA damage and cancer

DNA damage is caused by the agents that can either damage one of its 3 billion bases or break the phosphodiester bond on which bases reside. From the perspective of cancer DNA damage: 1. causes the disease; as evidenced by many human-cancer susceptibility-syndrome arising from mutations in genes involved in DNA damage response (table 4), 2. used to treat the disease and 3. Responsible for the toxicity or side effects of therapies for the disease such as bone marrow suppression, gastrointestinal toxicities, and hair loss which causes DNA-damage-induced cell death of proliferating progenitor cells in these tissues [Kastan M.B et al., 2004].

Table 4: List of genes involved in cancer susceptibility syndrome

Disease	**Gene**	**Cancer predisposition**
Ataxiatelangiectasia (A-T)	ATM	Leukemia, lymphoma
Nijmegen breakage syndrome(NBS)	NBS1	Leukemia, lymphoma
A-T-like disorder (ATLD)	Mre11	Leukemia, lymphoma
Fanconi's anemia (FA)	FancD2, Brca2 (or FancD1)	Acute myelogenous Leukemia
Familial breast, ovarian carcinoma	Brca1, Brca2	Breast, ovarian, scattered others
Li-Fraumeni syndrome	p53, CHEK2	Sarcomas, leukaemias, brain tumors, adrenal tumors, others

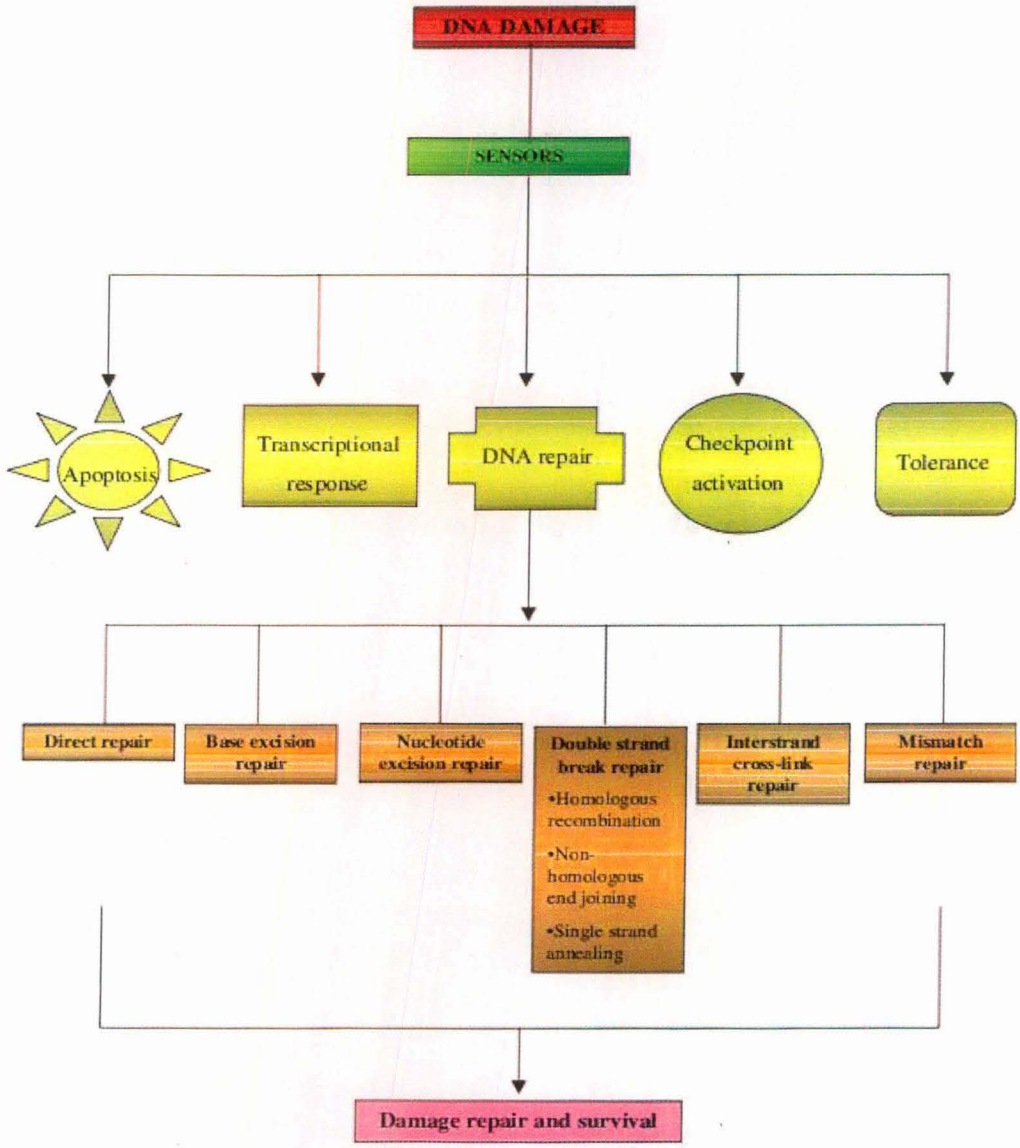


Fig 5: DNA damage responses related to cancer susceptibility in mammalian cells [Madhusudan S, 2005].

Cells have developed several elegant but not perfect mechanisms to cope with the constant endogenous and exogenous attack. A variety of different repair mechanisms exist for various types of DNA lesion that can occur (Fig 5). In addition to directly repair DNA breaks or adducts, cells respond to DNA damage by undergoing programmed cell death or by halting cell-cycle progression.

3.2 Cell cycle checkpoints and cancer

The term 'cell-cycle checkpoint' refers to mechanisms by which the cell actively halts progression through the cell cycle until it can ensure the completion of earlier process such as DNA replication and mitosis [Hartwell, 1989]. The checkpoint and repair pathways facilitate cellular responses to DNA damage, suggesting that DNA damage from both endogenous and exogenous sources is a major contributor to the development of human cancer. So, it is reasonable to speculate that alterations in these pathways increase the risk of cancer development [Fig 6; Kastan M.B et al., 2004].

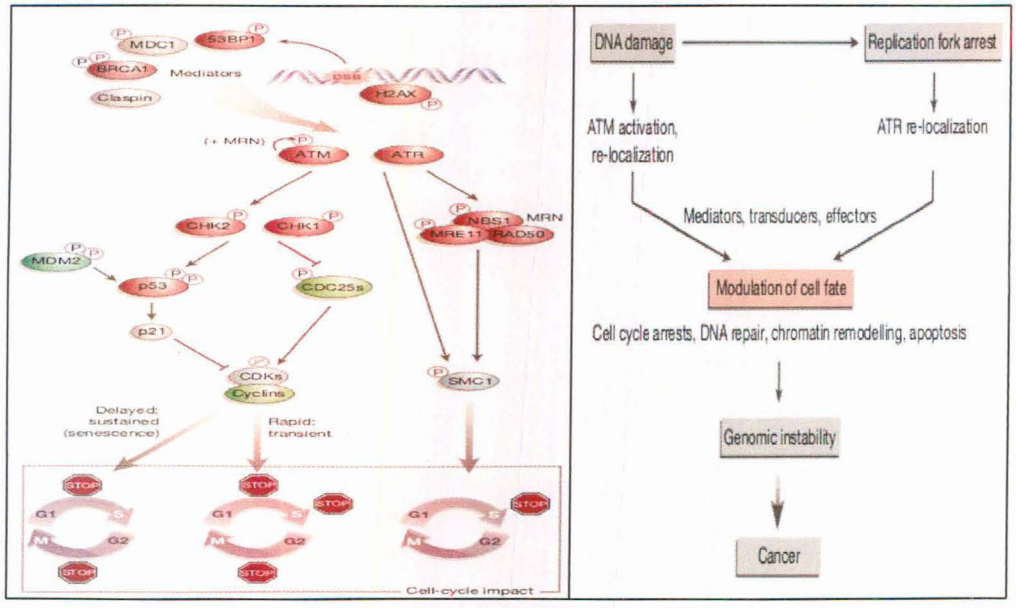


Fig 6: Cell cycle checkpoint pathway and cancer development (proto-oncogenes and tumor suppressors are shown in green & red respectively) [Kastan M.B et al., 2004].

The proximal checkpoint kinases ataxia telangiectasia mutated (ATM) and ataxia telangiectasia and Rad3 related (ATR) phosphorylate diverse components of the network, either directly or through the transducing kinases CHK2 and CHK1. The BRCA1 protein also contributes to cell-cycle arrest and DNA repair by homologous recombination, whereas p53 controls genes involved in cell death and DNA-repair mechanisms. Effectors pathways blocking cell cycle phases and p53 mediated permanent arrest (senescence), the global checkpoint network regulated by

ATM/ATR and CHK2/CHK1 affecting cellular responses other than cell cycle progression, including DNA repair, transcription, chromatin assembly, cell death and involvement of various proto-oncogenes (green) and tumor suppressors (red) need to be focused to understand the detailed mechanism of cancer progression (Fig 6).

Oncogenes

Oncogenes initially identified as genes carried by viruses cause transformation of their target cells. Major classes of viral oncogenes, which have their counterparts involved in normal cell functions, are called proto-oncogenes and their mutations or aberrant activation in the cell to form oncogene is associated with tumor formation. Oncogenes fall into several groups' i.e. transmembrane protein to transcription Factor. The generation of an oncogene represents a gain of function in which a cellular proto-oncogene is inappropriately activated. It involves a mutational change in protein or constitutive activation, over expression or failure to turn off expression at appropriate time. About 100 oncogenes have been identified. The oncogenes carried by DNA Viruses specify proteins that inactivate tumor suppressors, so their action in part mimic loss–of-function of tumor suppressors while those carried by Retroviruses are derived from cellular genes & mimic the behavior of Gain-of-Function mutation in animal proto-oncogene. Proto-oncogenes can be activated in several ways: i) Somatic mutation - exchange and loss/gain of bases ii) Amplification - formation of multiple copies of a gene which increases its expression iii) Activation of a gene by placement near a strong promoter -by invading viruses but in human tumors by reciprocal chromosome translocation. The known activations of proto-oncogenes in breast cancer are currently limited somatic mutations of ras genes and to the amplification of erb B_2 also known as HER-2 or neu gene [Callahan & Campbell, 1989]. Amplification of c-myc & int-2 genes are also considered by some to be responsible for the neoplastic mode of growth of breast cancer cells.

Tumor Suppressor Genes (TSG)

Some genes suppress tumor formation. Their protein product inhibits mitosis. When mutated, the mutant allele behaves as a recessive; that is, as long as the cell contains one normal allele, tumor suppression continues. (Oncogenes, by contrast, behave as dominants; one mutant or overly active allele can predispose the cell to tumor formation).

Example 1: RB - the retinoblastoma gene. Retinoblastoma is a cancerous tumor of the retina. The Rb protein prevents cells from entering S phase of the cell cycle. It does this by binding to a transcription factor called E2F. This prevents E2F from binding to the promoters of such proto-oncogenes as c-myc and c-fos. Transcription of c-myc and c-fos is needed for mitosis so blocking the transcription factor needed to turn on these genes prevents cell division. A random mutation of the remaining RB locus in any retinal cell completely removes the inhibition provided by the Rb protein, and the affected cell grows into a tumor (Knudson 1978). So, in this form of the disease, a germline mutation plus a somatic mutation of the second allele leads to the disease.

Example 2: p53 gene - The product of the tumor suppressor gene p53 is a protein of 53 kilodaltons. The p53 protein prevents a cell from completing the cell cycle if its DNA is damaged or the cell has suffered other types of damage. When the damage is minor, p53 halts the cell cycle, hence the cell division, until the damage is repaired. If the damage is major and cannot be repaired, p53 triggers the cell to commit suicide by apoptosis. This function makes p53 a key player in protecting against cancer and thus is an important tumor suppressor gene. More than half of all human cancers, harbor p53 mutations and have no functional p53 protein.

3.3 Types of Cancers

Cancers that occur on the skin or an organ are called carcinomas. Those that occur in muscle tissue or bone are called sarcomas. Lymphomas are tumors of the lymphatic system, affecting lymph nodes in the neck, groin, armpits, liver, and spleen. Leukemias are cancers of the blood-making bone marrow.

3.4 Breast Cancer

Breast Cancer is most often discovered in the form of a lump on the breasts which are benign, that is they are not cancerous or life threatening. When a tumor turns out to be malignant, however, it can be life threatening.

Breast cancer genetics. The inheritance of a single mutated allele of either BRCA1 or BRCA2 markedly increases the incidence of breast and ovarian cancers in women. As the tumors from these individuals virtually always lose the second allele, both

BRCA genes conform to the classic pattern of tumor suppressor genes [Venkitaraman A R et al., 2002].

About 10% of breast cancer cases cluster in families; some are due to highly penetrant germline mutations in one or another of a small number of genes, such as BRCA1 or BRCA2, giving rise to high cancer risk. But the majority of the cases (sporadic) exhibit no clear cut familial clustering and probably result from the collective effect of multiple, poorly penetrant variations in a much larger group of genes, modified by environmental factors. The familial case differs from the sporadic in having strong family history, onset at a young stage, involvement of tumors in the other organs like ovary, prostate etc. and bilateral occurrence. Of familial cases, germline mutation in BRCA1 or BRCA2 account for between 15% and 20% of the observed risk. Additional susceptibility alleles for familial breast cancer must therefore exist, but they have so far proven difficult to identify [Easton DF, 1999].

All breast epithelial cells of a BRCA mutation carrier have one inactivated allele of BRCA1 or BRCA2. During puberty, in response to estrogen surges, these cells rapidly proliferate. It is likely that this dramatic increase in the rate of cellular replication strains the DNA repair capacity of breast epithelial cells. Somatic genomic alterations involving the repeat elements of BRCA1 or BRCA2 will occur at high frequency.

Most cells that have both inherited and somatic inactivating mutations of BRCA1 or BRCA2 will be unable to repair DNA damage sustained in the following cell cycle and will die. However, in the rapidly proliferating breast epithelium, some repair deficient cells may escape death. Because these BRCA-null cells are deficient in repair, they would sustain DNA damage at many sites, often including genes essential to cell cycle checkpoint control. Mutation of a checkpoint gene would enable a BRCA-null cell to escape death permanently and to proliferate. Tumors in patients with germ line BRCA1 or BRCA2 mutations are frequently associated with somatic mutations of p53 (Welsch et al. 2001). Recent reports reveal that amplification of the MYB oncogene (Kauraniemi et al. 2000) and reduction of the anti-apoptotic gene Bcl-2 (Freneaux et al. 2000), are characteristic of most breast tumors from BRCA1 mutation carriers. On the other hand, somatic BRCA mutations rarely occur in sporadic cancer cases.

3.5 Targets for Cancer

Despite surgical treatment, irradiation and relatively new approaches of immunotherapy or vaccination, chemotherapy still remains an important option for treatment of malignancies. The success of chemotherapy is mainly hindered by identification of suitable target molecule or enzyme. To identify targets and molecules that select specifically tumor cells without causing life threatening infection and side effects is the ideal goal of successful chemotherapeutic treatment. To achieve this goal many pathways which hallmark cancer have continuously been explored to find significant difference between normal and malignant cells.

Potential targets of cancer [Szekeres T et al., 2002, Kim H J et al., 2006] which can be modulated by drugs widely used in cancer chemotherapy as well as antiviral therapy can be broadly classified in to: 1. DNA Methylation [Szyf M, 2000] 2. Caspade proteinase- Caspases [Henkel K.M et al., 1999] 3. Various protein Kinases [Longati P, 2001] 4. Metalloproteinases [Lee KS, 1996] 5. Topoisomerases I & II [Huang CH, 2001] 6. Farnasyl Transferase [Johnston S.R.D, 2001] 7. Telomerase [Lavelle F, 2000] 8. Prolactins [Llovera M, 2000] 9. Cell Surface Antigen [Wick B, 1997] 10. Signal transducer & Activator of Transcription [Turkson J, 2000] 11. Other enzymes related to protein phosphorylation [Hanover JA, 2001] 12. Ribonucleotide Reductase [Elford H.L, 1970] 13. Use of Hypoxia-selective Cytotoxins [Brown J.M 1998, 1999].

Screensaver Project which was launched in April, 2001 by Center for Computational Drug Discovery addressed the cancer target proteins listed in table 5 each of which has been tried against 3.5 billion molecules against its established active site [Richard W.G, 2002].

Table 5: Cancer target proteins addressed in Screen saver project (April, 2001).

PDB code	Name	Comment
821P	RAS	Signals cell growth
1FLT	VEGF	Growth of blood vessels
1ISC	SOD	Scavenger of superoxide radicals (for example, in leukaemia) that otherwise cause apoptosis
1IR3	Insulin-receptor tyrosine kinase	Representative of certain signalling tyrosine kinases
6COX	COX-2	Increases blood supply
1IEP	BCR–ABL (a tyrosine kinase)	Believed to be causally involved in chronic myelogenous leukaemia
1AGW	FGFR	Increases blood supply
1AQ1	CDK2	Regulates cell cycle
1C1Y	RAF	Receptor for activated RAS protein
1D8D	FPT	Responsible for activating RAS
1BZH	PTP1B	Regulates certain cellular events
1FLT	VEGFR1	One of the receptors for VEGF

COX-2, cyclooxygenase-2; CDK2, cyclin-dependent kinase 2; FPT, farnesyl protein transferase;FGFR, fibroblast-growth-factor receptor; PDB, Protein Data Bank; PTP1B; protein tyrosine phosphatase 1B;SOD, superoxide dismutase; VEGF, vascular-endothelial growth factor; VEGFR, VEGF receptor.

3.6 BRCA1 -a target protein

BRCA gene products (BRCA1 & BRCA2) participate in cellular responses to DNA damage, but they seem to have distinct roles. BRCA1 is a target of the ATM, ATR and CHK2 Kinases and is required for cell-cycle checkpoint responses in S phase and G2/M (Xu B et al., 2001). BRCA1 also localizes to sites of DNA breakage, interacts with chromatin remodeling proteins and has been implicated in transcription control [Venkitaraman A R, 2002]. The BRCA proteins function in pathways that repair DNA double-strand breaks (DSBs), stalled replication forks and DNA cross links.

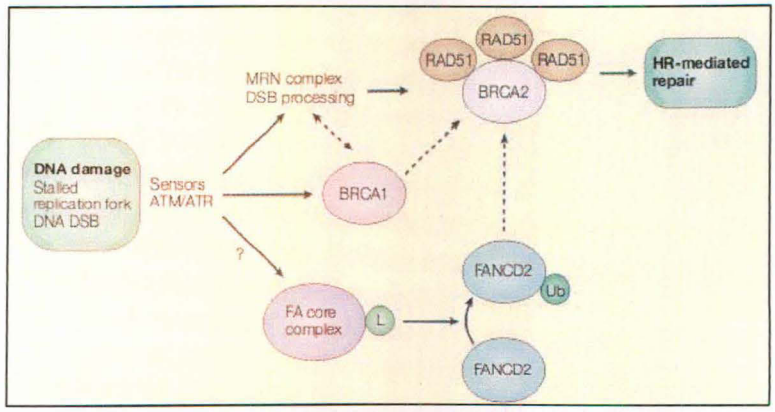


Fig 7: BRCA DNA repair pathway [Turner N, 2004]

DNA damage is sensed by protein kinases such as ataxia telangiectasia and Rad3 related (ATR), and ataxia telangiectasia mutated (ATM) that activates the pathways (Fig 7). Five FA (Fonconi Anemia) proteins i.e. FANCA, FANCC, FANCE, FANCF and FANCG form a nuclear complex5, which interacts with FANCL in response to DNA damage leading to mono-ubiquitylation of FANCD2. Ubiquitylated FANCD2 subsequently co-localizes with both BRCA1 and BRCA2 in nuclear foci. Homozygous mutations in BRCA2 also cause FA (Fonconi Anemia); BRCA2 is also known as FANCD1. BRCA2 regulates the RAD51 recombinase that mediates strand invasion and homology directed repair [Turner N, 2004].

These suggested functions of BRCA gene product is critical for tumor suppression and in order to get the relative specificity for breast and ovarian cancers associated with its mutation, BRCA associated proteins structure function has been studied here.

3.7 New approaches to molecular cancer therapeutics

Burchenal and his colleagues used analogs of folic acid, methotrexate, and of a nucleoside, 6-mercaptopurine, to induce profound and sustained remissions in children with aggressive leukemias [Burchenal, 1956] which proved revolutionary as for the first time (fifty years ago), drugs of known chemical composition that interfered with enzymes engaged in a specific biological process, DNA replication, were used to treat cancers successfully in a rational manner. [Varmus H, 2006].

Now more than 350 cancer genes have been catalogued (www. sanger. uk.ac/ Genetics/ CGP/census/). Genetic and epigenetic changes like DNA methylation, Histone modification [Esteller M, '07] and new approaches like RNAi platforms have recently been focused for new target validation, selection and prioritization. Despite decades of academic cancer research and investment in genomics, genetics and automation; many potential cancer targets remain undrugged (e.g. P53, RAS, MYC and oncogenic pathway WNT). Only 5% of cancer drugs entering clinical trials reach marketing. Success of cancer drug development is exemplified by high-profile drugs such as Herceptin (trastuzumab), Gleevec (imatinib), Tarceva (erlotinib) and Avastin (bevacizumab). Now, cancer drug development is leading the way in exploiting molecular biological and genetic information to develop 'personalized' medicine. [Collins I, 2006]. Greater emphasis on achieving therapies by exploiting DNA repair abnormalities such as BRCA mutants is justified [Farmer H, 2005].

CHAPTER 4.

BARD1- BRCT domain: A Case Study

4.1 BARD1: a unique protein

Structure of BARD1 The BRCA1-associated Ring Domain (BARD1) protein was discovered in a yeast-two hybrid screen as a binding partner of BRCA1. BARD1 structure is very similar to BRCA1 (Fig 8): 1. both proteins have a RING (green) domain which mediates DNA–protein and protein–protein interactions, 2. a nuclear export signal (NES, brown) at their N termini and 3. two tandem BRCA1 carboxy-terminal (BRCT, red) domains.

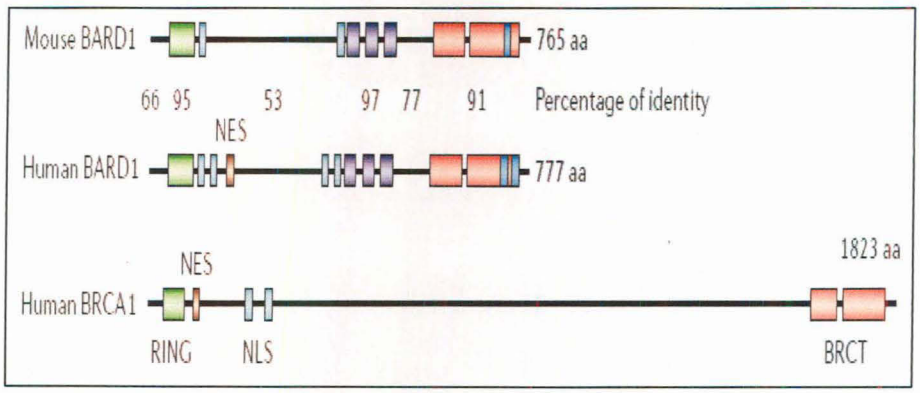


Fig 8: BARD1 domain compared to BRCA1 [Irminger-Finger I, 2006]

BARD1 and BRCA1 have no sequence or structural similarity with BRCA2. In addition to RING and BRCT domains, BARD1 has three ankyrin repeats (ANK, blue) that also facilitate protein–protein interactions. Interestingly, no other proteins that contain RING, ANK and BRCT domains were identified in a database search using the Simple Modular Architecture Research Tool (SMART).

BARD1-BRCA1: BARD1 and BRCA1 form a functional heterodimer through the binding of their RING-finger domains (PDB id 1JM7) [Brzovic P S et al., 2001] (Fig 9), which is thought to stabilize both proteins, as the respective monomers are unstable [Meza et al.,1999, Joukov et al., 2001].

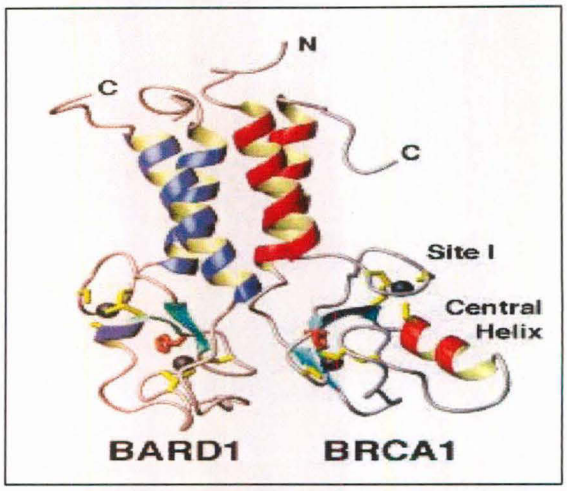


Fig 9: Crystal structure of BRCA1-BARD1 Heterodimer (PDB id 1JM7) [Brzovic P S et al., 2001]

BRCA1–BARD1 interaction has ubiquitin ligase activity [Hashizume R, 2001], play critical role in genomic stability through its function in cell cycle check point control and DNA repair. Furthermore, the interaction is required for several of the cellular and tumor-suppressor functions of BRCA1, as specific mutations within the BRCA1 RING domain are associated with breast and ovarian tumor. BARD functions are shown in Fig 10 [Irminger-finger et al., 2002].

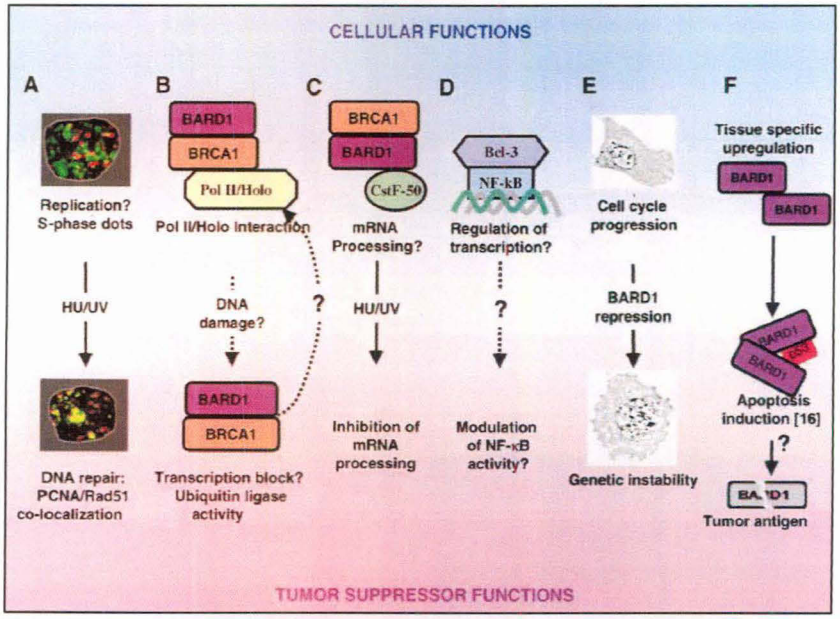


Fig 10: Compilation of BARD functions (vertical arrow shows transition from cellular to tumor suppressor function) [Irminger-finger et al., 2002]

BARD1 was interpreted only as accessory protein for BRCA1, since few cancer associated mutation have been reported in BARD1 than BRCA1 (more than 650), reports demonstrating BRCA1- independent function of BARD1, primarily in apoptosis and BRCA1-independent increased expression of BARD1 during mitosis suggest crucial independent function of BARD1.

BARD1 orthologues: BARD1 orthologues have been identified in mouse, rat, *Xenopus laevis* [Joukov V et al., 2001], *Caenorhabditis elegans* [Boulton S J et al., 2004], and *Arabidopsis thaliana* [Lafarge S et al., 2003]. The domain structure and the intron-exon boundaries that lie within the regions that encode the RING domain, and the region that includes the ANK repeats through to the BRCT domains, are conserved between the species. In addition to RING, ANK and BRCT domains, protein-sequence analysis predicts that BARD1 might possess an uncharacterized domain (I.I.-F., unpublished observations as reviewed in Irminger-Finger I, Nature, 2006) that is located between the ANK and BRCT domains, which are also highly conserved in various species. This complexity of structure indicates that BARD1 could have multiple functions. These might be regulated by the post-translational modifications (sites indicated by P; Fig 11) and expression of differentially spliced isoforms (BARD1 β and γ expressed by preleptotene spermatocytes whereas BARD1 δ is over expressed in ovarian cell line [Feki A et al., 2004] and in HeLa cells [Tsuzuky M et al., 2005]-in ovarian cancer samples the N-terminal exons 2-6 are frequently lost [Wu j y et al., 2006]. (In- intron) (Fig 11).

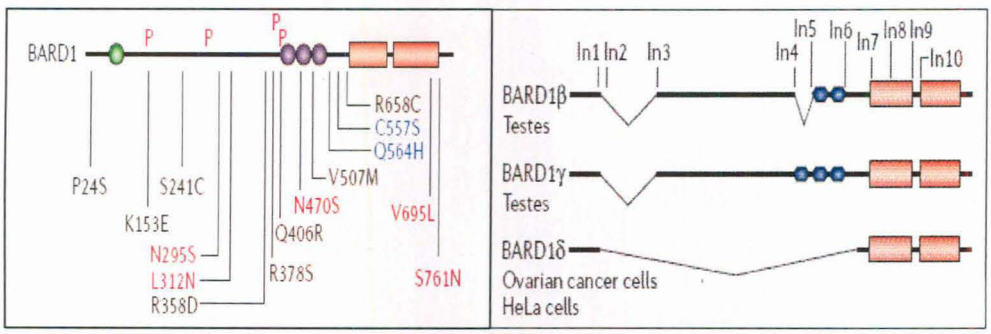


Fig 11: BARD1 domain; its phosphorylation & mutation site (left) and splice variants (right) [Irminger-Finger I, 2006].

BARD1 mutations and cancer: Because mutations in BRCA1 and BRCA2 account for only 50% of familial breast and ovarian cancer cases (according to data in the Human Gene Mutation Database), BARD1 mutations were expected to account for additional cases of inherited and sporadic breast and ovarian cancers. Screening of patients with sporadic breast, ovarian and endometrial cancers identified three missense alterations at amino-acid positions Q564H, V695L and S761N, and loss of heterozygosity (LOH) [Thai T H et al.,1998] was associated with Q564H and S671N mutations (mutations- red, germline mutation- blue, polymorphism-black) (Fig 11).

4.2 Functional domain of BARD1

Human BARD1 is a 777 amino acid protein that has a number of interacting regions: a RING finger (residues 46-90; green), three ankyrin (ANK; residues 420-525; blue) repeats and two tandem BRCT (residues 568-777; red) domain. Besides, location of potential nuclear localization signal (NLS, light blue), position of phosphorylated serine or threonine (P), phosphorylation that competes for BRCA1 binding and ubiquitin ligase functions (P*), region required for homodimer formation and ubiquitin ligase activity are at N terminus region. Regions at C terminus are - the minimal apoptotic region, F1 immunogenic regions (Gautier F, 2000) and phosphobinding regions are shown in Fig 12.

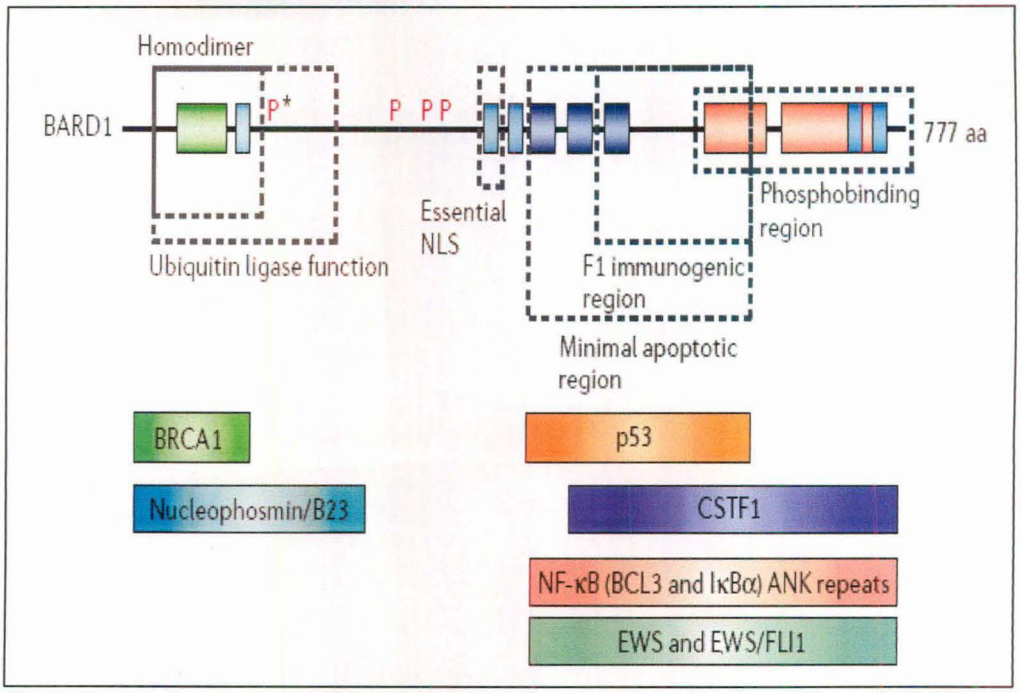


Fig 12: Protein interactions and functional domains of BARD1 [Irminger-Finger I, 2006]

These regions have been shown to interact with a number of proteins (bars) e.g. BARD1, CSTF1, cleavage stimulating factor subunit 150KDa [Kleiman, 2001; Edwards, 2008] the Ewing sarcoma gene product (EWS) and the oncogenic fusion protein EWS–FLI1 [Spahn L et al., 2002] which is consistent with an oncogenic function of BARD1 (alternatively, EWS might sequester BARD1, therefore inhibiting its tumor-suppressor functions), 1κBa, NF-κB inhibitor, alpha.

BARD1 Pathways: BARD1 participates in two major pathways (Fig 13). The first is (a) a cell survival pathway mediated by the BRCA1-BARD1 heterodimer. The second is (b) a cell death pathway, which is independent of BRCA1.

In pathway (a), the activity of the BRCA1-BARD1 ubiquitin ligase leads to RNA pol II degradation and cell-cycle arrest, to γ tubulin degradation and control of centrosome duplication, to H2A/H2AX ubiquitylation and epigenetic control, and to Nucleophosmin (NPM) ubiquitilation and stabilization. Increased expression of NPM is known to inhibit apoptosis, and it causes centrosome amplification and genomic instability, hence antagonizes BARD1 functions.

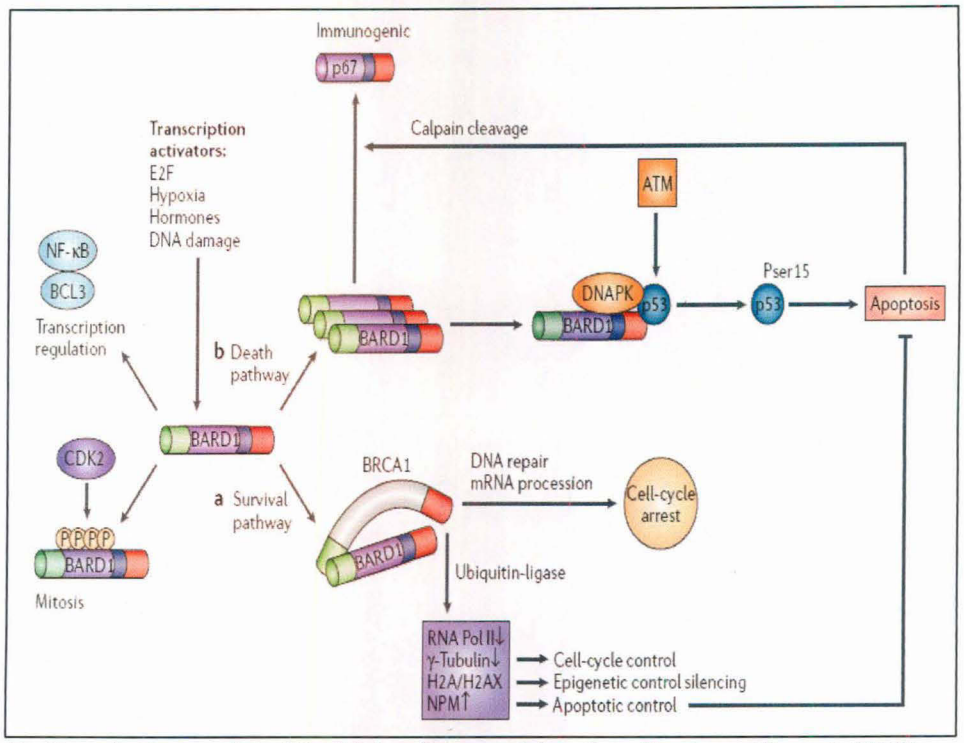


Fig 13: Hypothetical model of BARD1 pathway and functions [Irminger-Finger I, 2006]

In pathway (b), expression of BARD1 can be increased by DNA damage, exposure to ultraviolet light, hypoxia and hormone signaling. Increased expression of BARD1 stabilizes p53 and facilitates its phosphorylation by DNA-dependent protein kinase (DNAPK). The role of BARD1 in p53 phosphorylation at serine 15 (pSer15) by ATM is not known. Post-translational modification, through phosphorylation by CDK-cyclin complexes, might regulate the interaction of BARD1 with BRCA1 and trigger its mitotic activity. BARD1 also has transcriptional activity as it can induce transcription activity of NF-κBs through binding to the NF-κB co-factor BCL3 (B cell leukemia/lymphoma3). Finally, the proteolytic cleavage product of BARD1 (p67) is immunogenic and has anti-tumorigenic properties.

4.3 BRCT domain

Initially identified as a sequence motif homologous to the BRCA1 COOH-terminal region (Fig 14), the BRCT domain is a protein domain approximately 90 amino acids in length [Wu, L. C., 1996].

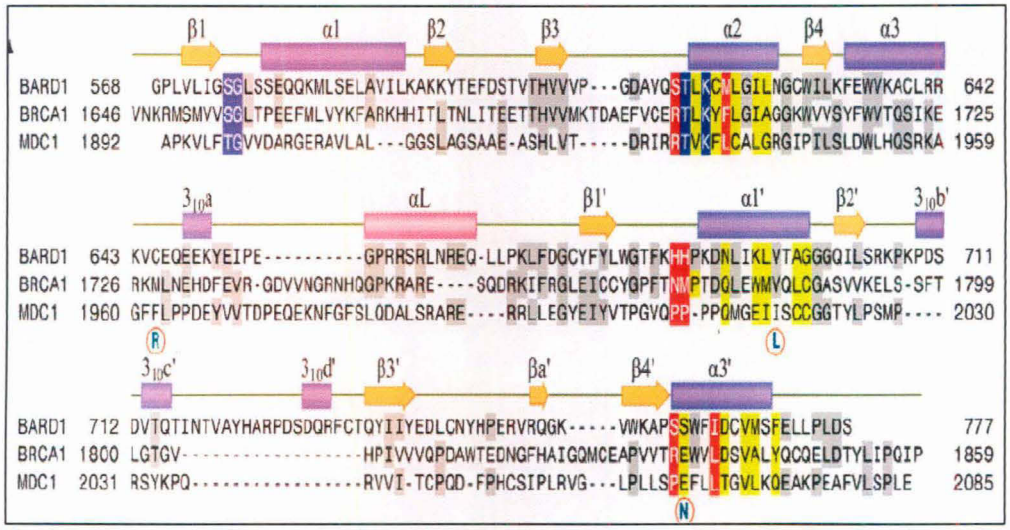


Fig 14: Sequence alignment of the BARD1, BRCA1 & MDC1 BRCT domains using CLUSTAL W [Birrane G, 2007]

It is an ancient protein module that can be found in single cell eukaryotes [Aravind L et al., 1999]. In the human genome more than 30 BRCT containing proteins have been documented. A number of proteins playing key role in DNA-damage checkpoint control and DNA repair has BRCT repeats which has phosphopeptide-binding modules (e.g. BRCA1, MDC1, BARD1, and DNA Ligase IV) [Glover JNM et al., 2004, Manke et al., 2003, Rodriguez et al., 2003]. Heterodimerization between single BRCT domains (*e.g.* XRCC1 and DNA Ligase III) has been reported [Taylor, R. M, 1998]. In addition, BRCT domains of 53BP1 can interact with P53, while BRCT domains of BRCA1 bind DEAH family helicase BACH1 (BRCA1-associated C-terminal helicase) and CtBP interacting protein CTIP (Chai, Y, 1999, Yu, X., 1998). The BRCT domains might also bind double stranded DNA breaks (Yamane, K et al., 1999). List BRCT-mediated protein–protein interactions are shown in table 6.

Table 6. BRCT-mediated protein–protein interactions*

BRCT–BRCT interactions		
BRCTcontaining protein	**BRCT containing binding partners**	**Cellular function**
XRCC1(N-terminal)	PARP	Base excision repair
XRCC1(C-terminal)	DNA ligase III	Base excision repair
ScRAD9	ScRAD9	DNA damage response
BRCT–non-BRCT interactions BRCTcontaining		
BRCTcontaining protein	**Binding partners**	**Cellular function**
BRCA1	BACH1	G2/M checkpoint
BRCA1	CtIP and LMO4	Transcriptional regulation
BRCA1	HDACs	Chromatin structure
BRCA1	CBP	Chromatin structure
BRCA1	P53	Transcription
BRCA1	RNAP holoenzyme (RHA subunit)	Transcription
BRCA1	Acetyl-CoA Carboxylase	Fatty acid Metabolism
ScDNA ligase IV	LIF1	NHEJ
DNA ligase IV	XRCC4	NHEJ
TOPBP1(similar to SpCut5/Rad4,ScDpb11, DmMus101)	E2F1	DNA repair/checkpoint/ Transcription Repression
Rad4	Rad9(PCNA-like)	Checkpoint
Dpb11	Ddc1 (PCNA-like)	Checkpoint
Dpb11	Drc1	S-phase checkpoint
53BP1	p53	DNA repair/transcription/ checkpoint
SpCrb2 (similar to 53BP1)	Rad3 kinase	Checkpoint
FCP1	RNAPII CTD	Transcription
Swift	Smad2	Transcription

*Abbreviations: 53BP1: p53 binding protein 1; BACH1: BRCA1-associated C-terminal helicase; BRCA1: breast-cancer-associated protein 1; BRCT: BRCA1C-terminal; CBP: CREB-binding protein; CTD, C-terminal heptad repeat tail; Dm, Drosophila melanogaster; HDAC: histone deacetylase; NHEJ: non-homologous end joining; PARP: poly(ADP-ribose) polymerase; PCNA, proliferating-cell nuclear antigen; RNAP, RNA polymerase; Sc, Sacccharomyces cerevisiae; Sp, Schizosaccharomyces pombe; TOPBP1; topoisomerase IIb binding protein 1.

Recent crystallographic studies of BRCA1 and MDC1 BRCT domain bound to target peptides provide important clue to ligand recognition by these modules. [Shiozaki E.N et al., 2004, Clapperton J.A et al., 2004, Williams et al., 2004, Verma et al., 2005, Stucki M et al., 2005] (Table 7).

Table 7: List of Crystal structure of BRCA/BARD1, its bound ligands & mutants.

PDB ID	TITLE	REFERENCES
1JM7	Structure of a BRCA1-BARD1 heterodimeric RING-RING complex	Brzovic P.S et al., 2001
1JNX	Crystal structure of the BRCT repeat region from the breast cancer-associated protein BRCA1	Williams R.S et al., 2001
1N5O	Structural consequences of a cancer-causing BRCA1-BRCT missense mutation	Williams R.S et al., 2003
1OQA	Solution structure of the BRCT-c domain from human BRCA1	Gaiser O.J et al., 2004
1T29	Crystal structure of the BRCA1 BRCT repeats bound to a phosphorylated BACH1 peptide	Shiozaki E.N et al., 2004
1T2U	Structural basis of phosphopeptide recognition by the BRCT domain of BRCA1: structure of BRCA1 missense variant V1809F	Williams R.S et al., 2004
1T2V	Structural basis of phosphopeptide recognition by the BRCT domain of BRCA1	Williams R.S et al., 2004
1Y98	Structure of the BRCT repeats of BRCA1 bound to a CtIP phosphopeptide	Verma A.K et al., 2005
2NTE	Crystal structure of the BARD BRCT Domains	Birrane G et al., 2007
1ING	X-ray Structure of the BRCA1 BRCT mutant M1775K	Tischkowitz M et al., 2008
3COJ	Crystal Structure of the BRCT Domains of Human BRCA1 in Complex with a Phosphorylated Peptide from Human Acetyl-CoA Carboxylase 1	Shen Y, Tong LP, 2008

BRCA1-Ctlp complex (PDB entry 1Y98)

As shown in Fig 15; a shallow pocket P1 in N-terminal BRCT repeats receives the phosphoserine at position 0 of the ligand. It show the part of the active site from BRCA1-Ctlp complex (PDB entry 1Y98); Ctlp peptide (pink) and the unbound BARD1 (grey), interacting BRCA1 residues (green), water molecule (cyan sphere), whereas a deeper hydrophobic pocket (P2) in the groove between the N- and C-terminal BRCT repeats select residue at position +3 (Fig 16). The similarity of the amino acid composition and architecture of P1 pocket of BRCTs provides consistency by its nondiscriminatory binding to pser0 of the ligand. In contrast, the P2 pockets from various BRCT domain varies significantly that determine their selectivity properties [Birrane G, et al., 2007].

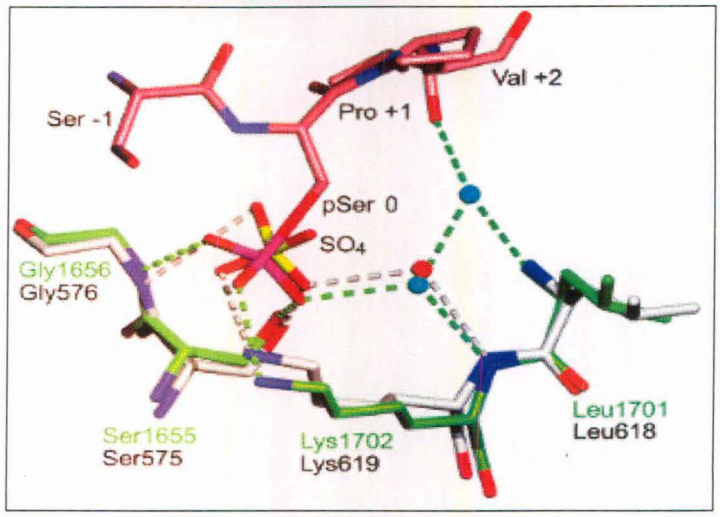


Fig15. Stick model superposition of the P1 pockets of the BRCA1 (green)-Ct1p (pink) complex (PDB entry 1Y98) and the unbound BARD1 (grey). Water molecule interacting with BRCA1 & BARD1 (Cyan & red sphere; respectively), Nitrogen, Oxygen, Phosphorus and sulphur atoms are shown in blue, red, magenta, yellow respectively [Birrane G, et al., 2007].

BRCA1 BRCT-Phosphopeptide complex (pdb id: 1T2V)

As shown in Fig 16, (a) an overview of the complex formed between the BRCT domain and phospho-peptide, highlighting the side chains that line the phospho-serine and phenylalanine (C3 position) binding pockets (pdb id: 1T2V). (b) Detail of phospho-peptide recognition. The optimized phospho-peptide is shown in blue, the BRCT domain of BRCA1 is shown in grey, and hydrogen bonding and salt bridges are indicated by yellow dashes. Alignment of the structure of the BRCT domain of 53BP1 (shown in green) with that of BRCA1 reveals structural conservation of the peptide-binding interface [Glover JNM et al., 2004, Williams R.S et al., 2004].

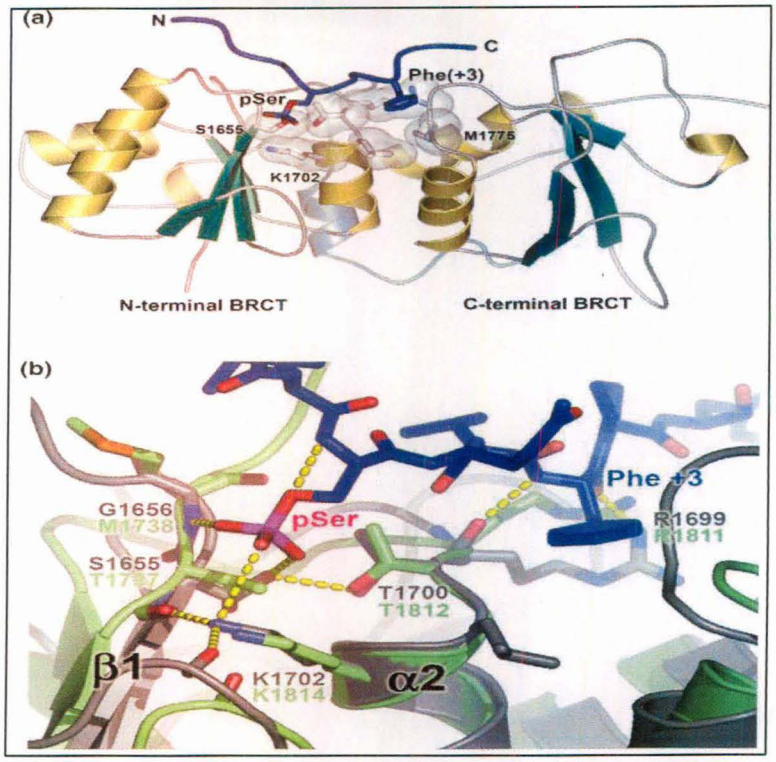


Fig 16: Phospho-peptide recognition by the BRCT domain of BRCA1 [Williams et al., 2004]

BRCA1-BACHI complex (PDB id: 1T29)

BRCT Repeats of Human BRCA1 bound to a Phosphorylated BACH1 (BRCA1-associated C-terminal helicase) Peptide is shown in Fig 17 (A, B).

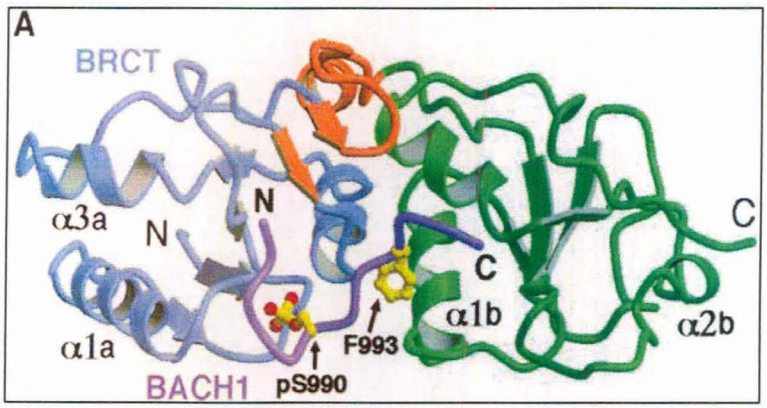
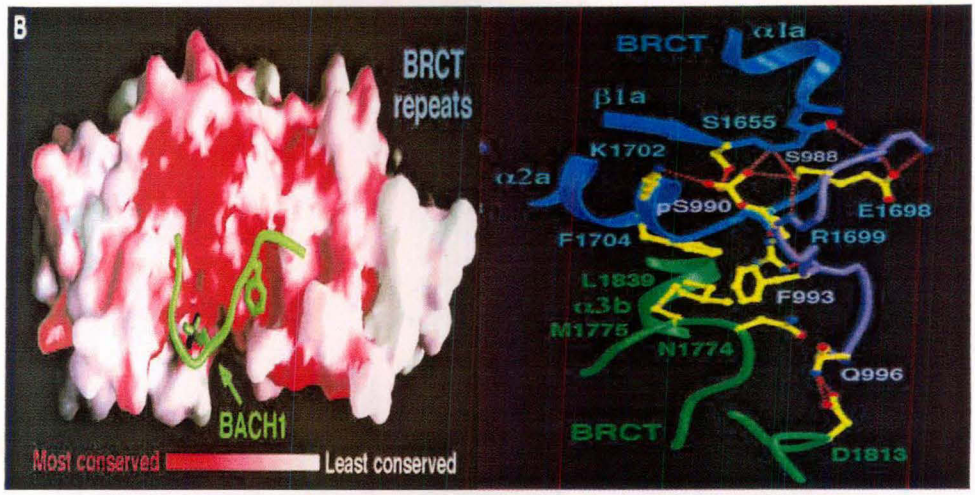


Fig 17 A: Structure of the BRCT Repeats of Human BRCA1 Bound to a Phosphorylated BACH1 Peptide. The two tandem BRCT repeats (blue and green), structural elements between the two BRCT repeats (orange), bound BACH1 peptide (magenta), Phe993 and the phosphorylated Ser990 of the BACH1 peptide are highlighted [Shiozaki et al., 2004]



In Fig17: (B) The phosphorylated BACH1 peptide binds to a highly conserved surface cleft of BRCA1 proteins across species (human, rat, chicken, and fish). In surface representation of the BRCT repeats; residues that are invariant in all four BRCA1 orthologs (red) while residues that is completely divergent (white) are shown. The bound BACH1 peptide is shown in green (left). Stereo view of the BRCT-BACH1 interface. The two tandem BRCT repeats are colored blue and green. The bound BACH1 peptide is shown in magenta, Hydrogen bonds; red dashed lines, Oxygen and nitrogen atoms; red and blue balls respectively are shown (Right) [Shiozaki E.N et al., 2004].

In Fig 17: (C) Protein-peptide contacts between tandem BRCT domains and the BACH1 phosphopeptide, (D) BRCT cancer-linked mutations (D1692Y, C1697R, R1699Q, S1715R, M1775R, Y1853X (red) & S1655F (green; as its cancer predisposition has not been confirmed by pedigree analysis) which cluster near the phosphopeptide-binding site and sequence conservation in relation to the BACH1 phosphopeptide-binding site are shown [Clapperton J.A et al., 2004].

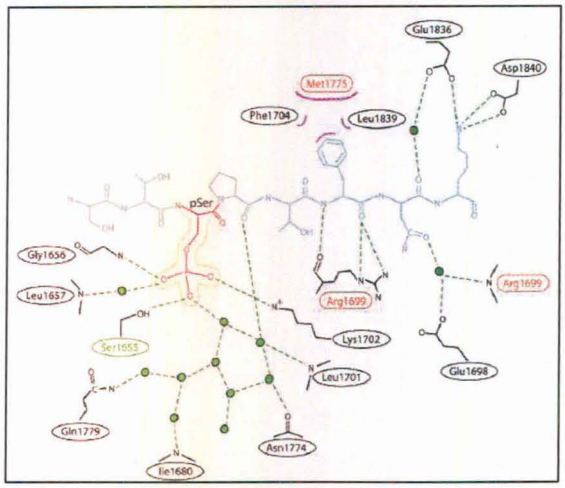


Fig 17 C: Protein-peptide contacts between tandem BRCT domains and the BACH1 phosphopeptide are shown. Dashed lines, hydrogen bonds; pink crescents, van der Waals interactions; green circles, water molecules. [Clapperton J.A et al., 2004]

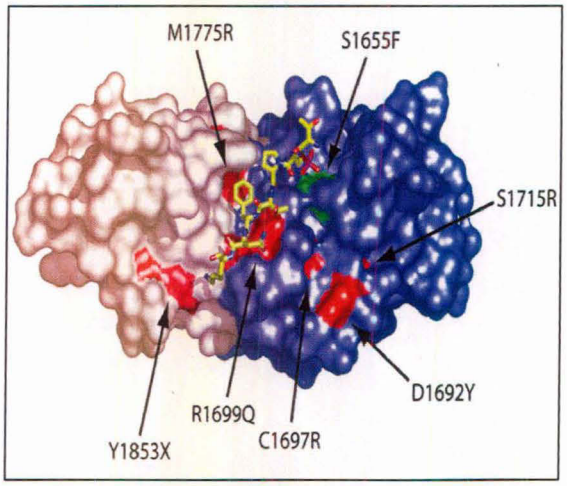


Fig 17 D: Molecular surface representation of tandem BRCT domains with BRCT cancer-linked mutations and sequence conservation in relation to the BACH1 phosphopeptide-binding site. BRCT1 (blue), BRCT2 (gray), and the mutations (red/green) are shown [Clapperton J.A et al., 2004].

Fig 17 E: The Phe (+3) position of the BACH1 phosphopeptide essential for tandem BRCT domain binding specificity is highlighted. (a) Phe1704, Met1775 and Leu1704 from tandem BRCT domains form a hydrophobic pocket to accommodate the Phe (+3) position of the BACH1 phosphopeptide. (b) Superposition of the crystal structure of the M1775R tandem BRCT domain mutant with the wild-type BACH1–

phosphopeptide complex reveals that this mutation occludes the BACH1 Phe(+3) position.

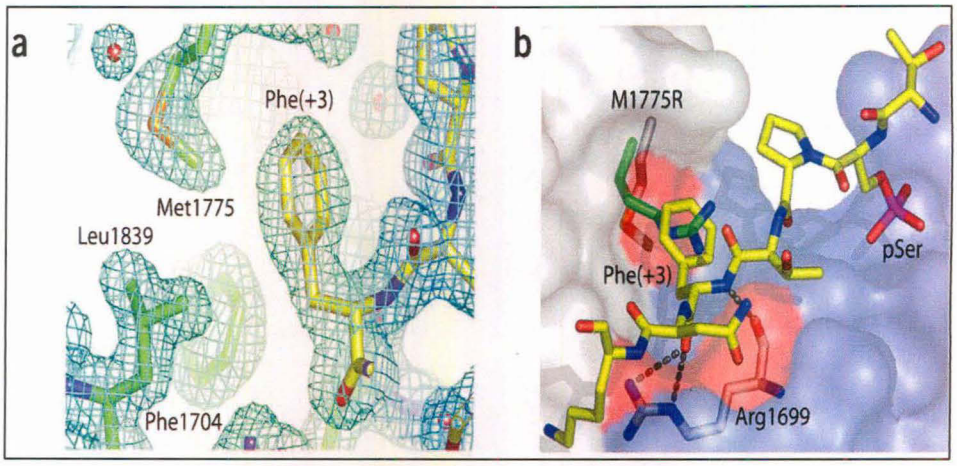


Fig 17 E: The Phe (+3) position of the BACH1 phosphopeptide (a). Crystal structure of BRCT mutant bound to wild type BACHI peptide (b) [Clapperton J.A et al., 2004].

MDC1 BRCT- γH2AX Tail Complex

The 207 residue C-terminal fragment of MDC1 retains the typical tandem BRCT fold (Williams et al., 2001) in which each BRCT repeat (BRCT 1 and BRCT 2) adopts a compact a/b fold and is connected by a linker region (Fig 18A). γH2AX binds in an extended conformation to a groove at the interface between the two BRCT repeats. Structure and sequence analysis of the MDC1 family reveals that the γH2AX binding groove has been highly conserved throughout evolution. Overall, the structure of the MDC1 BRCT- γH2AX complex explains the requirement for phosphorylation of H2AX Ser139, the overall sequence specificity, and the importance of a free C terminus [Stuki M et al., 2005] (Fig 18A, B).

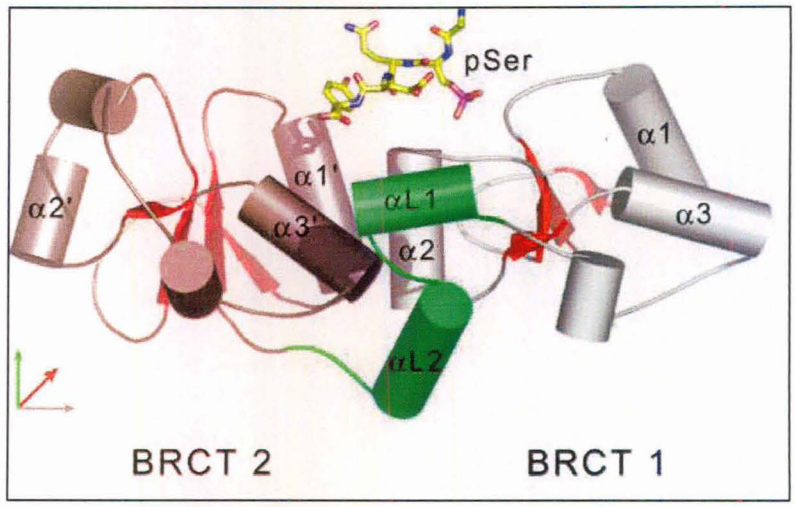


Fig 18 A: Ribbon representation of the MDC1 BRCT-γH2AX tail complex. The γH2AX peptide (yellow stick model) binds at the interface between the two BRCT repeats. The BRCT repeat linker is colored green. [Stuki M et al., 2005]

In Fig 18 :(B) Protein-peptide contacts between MDC1 BRCT and the γH2AX tail is shown. (C) Detailed ball-and-stick superposition of the phosphate binding pockets from MDC1 BRCT-γH2AX and BRCA1 BRCT-BACH1 complexes. Phosphate binding pockets from MDC1 BRCT-γH2AX and BRCA1 BRCT-BACH1 complexes are shown. Water molecules are shown as red (MDC1) or white (BRCA1) spheres [Stuki M et al., 2005]

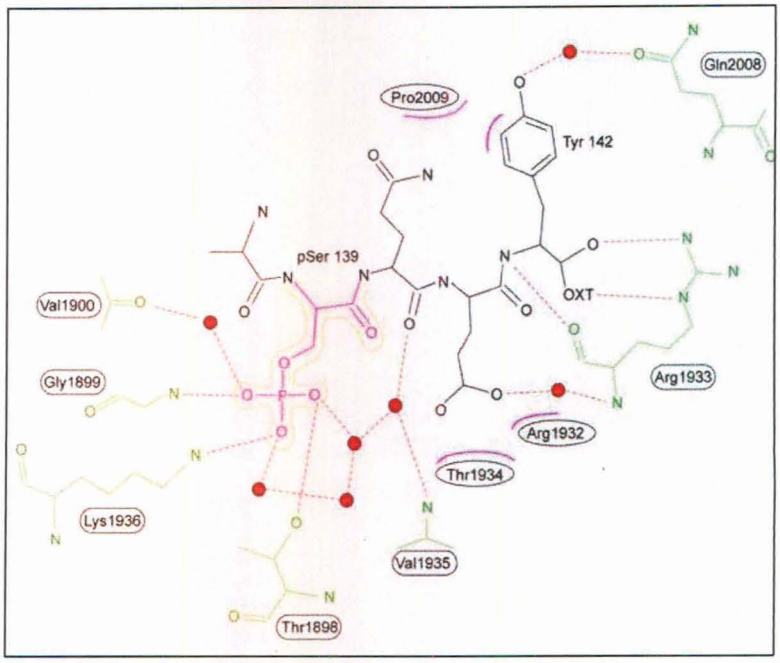


Fig 18 (B): Schematic representation of protein-peptide contacts between MDC1 BRCT and the γH2AX tail. Hydrogen bonds, van der Waals interactions, and water molecules are denoted by dashed lines, pink crescents, and red circles, respectively [Stuki M et al., 2005]

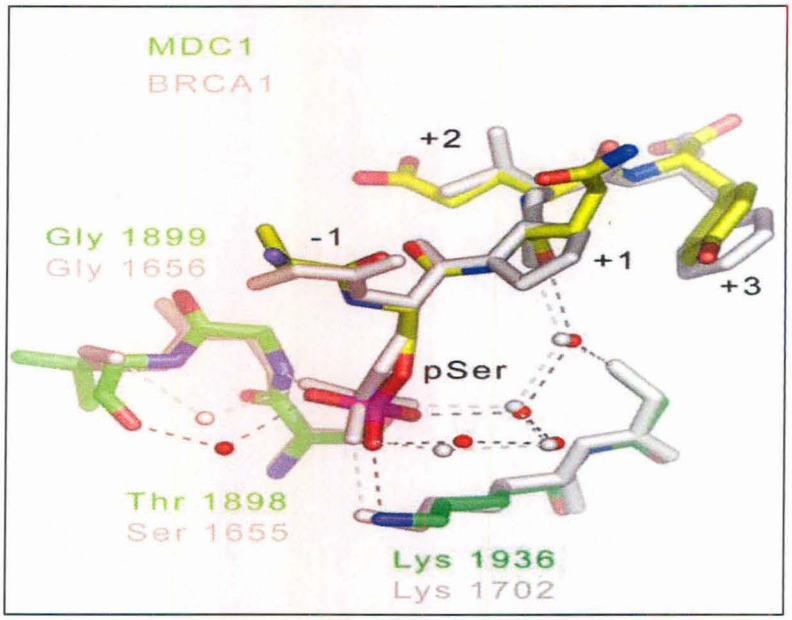


Fig 18 (C): Detailed ball-and-stick superposition of the phosphate binding pockets from MDC1 BRCT-γH2AX and BRCA1 BRCT-BACH1 complexes. Water molecules are shown as red (MDC1) or white (BRCA1) spheres [Stuki M et al., 2005].

4.4 Crystal Structure of BARD1- BRCT Repeat

Crystal structure of the human BARD1- BRCT repeats (residues 568-777) at 1.9 Å resolution (pdb id: 2NTE, Fig 19) shows composition and structure of the BARD1 phosphoserine-binding pocket P1 are strikingly similar to those of the BRCA1 and MDC1 BRCT domains, suggesting a similar mode of interaction with the phosphate group of the ligand. By contrast, the BARD1- BRCT selectivity pocket P2 exhibits distinct structural features, including two prominent Histidine residues, His685 and His686, which may be important for ligand binding. [Birrane G et al., 2007]

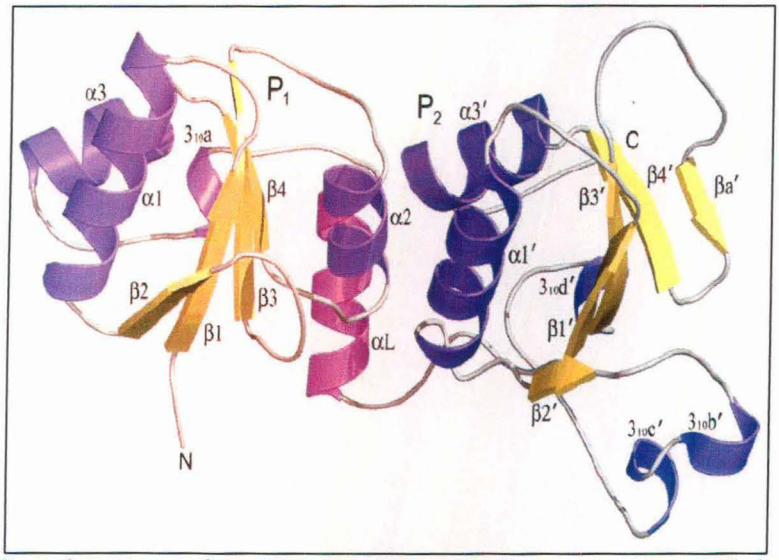


Fig 19: Crystal structure of BARD1- BRCT Domain at 1.9 angstrom resolution [Birrane G, 2007]

The region spanning residues 568-777 folds into two tandem domains, BRCT1 (residues 568- 654) and BRCT2 (residues 669-777) are linked by a central α-helix (αL). BRCT1 comprises a central β sheet formed by four parallel β-strands (β1- β4) and flanked by α-helices α1 and α3 on one side and α2 on the other. BRCT2 also consists of parallel β-sheet formed by the β strands β1'-β4', which are neighbored on one side by the α-helices α1' and α3'. However, the α2' helix that exists between β3' and β4' in other BRCT2 structure is replaced by a short antiparellel β-strand (βa') in BARD1 BRCT2 (Fig 19). Another unique feature of the BARD1 BRCTs is the presence of three short 3₁₀-helices (3₁₀b', 3₁₀c', and 3₁₀d') in the β2'-β3' loop.

The two BRCT modules pack closely against each other, burying a hydrophobic interface of ~1450 $Å^2$ and creating a surface groove with two pockets, P1 and P2. The hydrophilic P1 is formed by residues Ser575, Gly576, Thr617, and Lys619 (Fig15), conforming to the consensus motif (Ser/Thr)-Gly...Thr-X-Lys that is characteristic of the pSer interacting BRCA1 and MDC1 P1 pockets (Fig 15, 18C respectively). A sulfate ion from the crystallization solution is present in the P1 pocket, where it makes direct and solvent-mediated interactions with BRCT residues, specifically, the sulfate ion hydrogen bonds with the main chain N of Gly576, the O^{γ} of Ser575 and the N^{ζ} of Lys619 (Fig 15). These interactions are similar to those stabilizing the phosphate group of the ligand pSer0 in the BRCA1 and the MDC1 structures [Shiozaki et al., 2004, clapperton et al., 2004, Williams et al., 2004, Verma et al., 2005, Stucki et al., 2005], suggesting that BARD1 P1 residues may also be involved in similar interactions with the phosphate group of the ligand. The deeper and more hydrophobic pocket P2 is lined by Ser616, Met621, His685, His686, and Ile764 (Fig20).

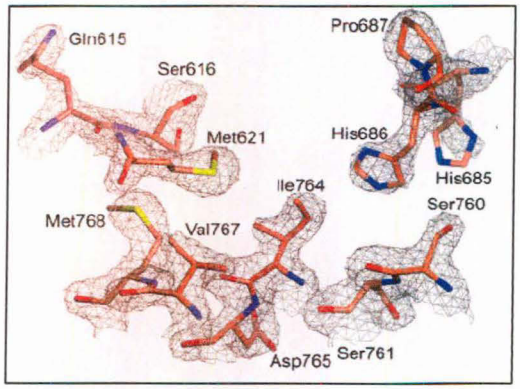


Fig 20: BARD1-BRCT P2 pocket with two prominent residues; His685 and His686

Two prominent features of this pocket are His685 and His686, corresponding to Asn1774 and Met1775 of BRCA1, respectively, that mediate interactions with Phe + 3 of the ligand [Shiozaki et al., 2004, Clapperton et al.,2004, Williams et al.,2004, Verma et al.,2005] suggesting that the BARD1 histidines could play a key role in ligand selection. Importantly, the BARD1- BRCT structure provides insights into the mechanisms by which the cancer associated missense mutations C645R, V695L, and S761N may adversely affect its structure and function.

Histidine Switch

The weak electron density of His685 imidazole group indicates the flexibility of its side chain while the imidazole group of His686 is well ordered and its position and geometry at the rim of P2 favors the formation of Hydrogen bond between its $N^{\varepsilon 2}$ atom and the side chain of a polar residue at position +3 of ligand. Importantly, the electrostatic potential of the P2 pocket is dramatically altered by changes in the protonation state of His685 and His686 (Fig 21). At near neutral pH, the solvent exposed $N^{\delta 1}$ and $N^{\varepsilon 2}$ atoms of His685 and His686 are not protonated, hence the P2 pocket has a negative electrostatic potential ,whereas at more acidic pH, these atoms gets protonated, switching the net charge to positive. This raises the possibility that the BARD1- BRCT interaction with its ligand/s may be dramatically regulated by the protonation of His685 and His686 in response to pH shift in local cellular environments during various physiological and pathological conditions [Boyer M J et al., 1992 & Lee, D et al., 2006]

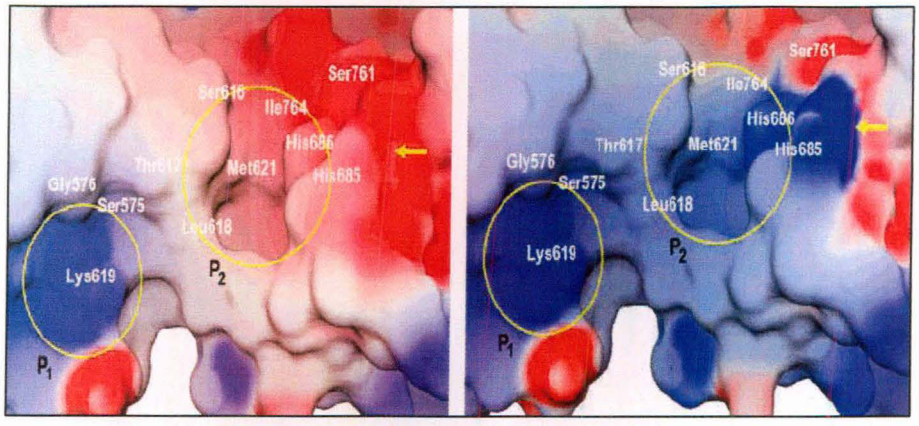


Fig 21: Electrostatic potential surface of BARD1- BRCT surface; at near neutral pH (left). at acidic pH (right). The effect of PH change is shown by color change of P2 pocket [Birrane G, 2007].

CHAPTER 5.

AIMS & OBJECTIVES

Against above background the work in this project explores the methods adopted for the successful and popular software used for vHTS. Hence the aims and objective is comparative study of different docking software with regard to the search algorithms and scoring functions. Initially T4 Lysozyme , a well characterized active site with 3D structure of ligand bound is studied for optimization of parameters using four docking software , Lysozyme co-crystallized with ligand isobutyl benzene (PDB ID: 184L.pdb) is chosen. Finally the target BARD1- BRCT domain (BRCA1 C Terminal domain), PDB ID 2NTE.pdb; well associated with tumor suppressor protein BRCA1 linked to breast and ovarian cancer is used to identify the active site(s) and then used for docking of ZINC database. Finally four methods are compared using ranking of the compounds as identified using each of the four docking protocols.

The software used for comparative study are as follows:

1. DOCK
2. FRED
3. GLIDE
4. GOLD

Target: BARD1- BRCT (BRCA1 C-terminal) Domain - 2NTE.PDB
Compound Library used: ZINC version 8 (http://zinc.docking.org)

CHAPTER 6.

MATERIAL & METHODS

6.1 Identification of Novel Cavities in BARD1-BRCT Domain

The knowledge of active site residue derived from literature or from visualization of the active site in addition to the ligand binding information if provided in the PDB structure, reduces both the error and CPU time required for docking. If the ligand bound structures are not available as in case of BARD1-BRCT domain (pdb id 2nte.pdb), identification of the cavity or expected space for ligand binding is crucial. Cavity detection programs may be used to identify the putative docking site. LigandFit [Venkatachalam C.M et al., 2003] is a novel method which in addition to docking ligands also has a program for cavity detection to identify and select the region of the protein as the active site for docking using the flood-filling algorithms to conveniently identify voids in the protein.

Table 8 contains number of programs for Protein Interface, ligand Binding Site, and Cavity Analyses from PDB [Kirchmair J et al., 2008]. Putative active site with spheres or PASS [Brady, 2000] employs simple geometric analysis to characterize regions of buried volumes in protein and identify binding sites based on size, shape and burial volume. The PASS algorithm is designed to fill the cavities in a protein structure with a set of spheres and to identify these spheres as active site points (ASPs) which represent the center point of the cavity predicted. Q Site Finder [Laurie, 2005] is a new method of ligand binding site prediction. It works by binding hydrophobic (CH3) probes to the protein, and finding clusters of probes with the most favorable binding energy. These clusters are placed in rank order of the likelihood of being a binding site according to the sum total binding energies for each cluster.

Table 8: List of programs to identify Protein cavity and interface.

Program from PDB	Role	References
PASS	identification of protein cavities	Brady G P et al., 2000
SURFNET	identification of protein cavities	Laskovsky, 1995
QSiteFinder	identification of protein cavities	Laurie, 2005
Protemot	prediction of protein binding sites with automatically extracted geometrical templates	Chang D T, 2006
SitesBase	comparative investigation of protein-ligand binding sites	Gold, 2006
PDBSITE	protein binding site analysis	Ivanisenko, 2005
PSIbase	repository of PDB-derived protein interface information	Gong,2005
PROTCOM	data pool of protein interface structures	Kundrotas,2007
Molsurfer	analysis and visualization of protein interfaces	Gabdoulline,2003
iPFAM	visualization and browsing of protein interfaces	Finns,2005
DMAPS	database of multiple alignments for protein structures	Guda,2006

6.1.1 Computed Atlas of Surface Topography of proteins (CASTp)

The updated CASTp web server (http://cast.engr.uic.edu/cast/) was used in this study to characterize surface features, functional regions and specific roles of key residues of proteins BARD1- BRCT domain (pdb id 2nte.pdb). This domain doesn't have any known bound ligand. Though the literature [Birrane G et al., 2007] suggests two characteristic pockets P1 and P2, a sulfate ion from the crystallization solution is present in the P1 pocket, where it makes direct and solvent-mediated interactions with BRCT residues. These interactions are similar to those stabilizing the phosphate group of the ligand pSer 0 in the BRCA1 and the MDC1 structures [Shiozaki et al.,2004, Clapperton et al.,2004, Williams et al.,2004, Verma et al.,2005, Stucki et al.,2005], suggesting that BARD1 P1 residues may also be involved in similar interactions with the phosphate group of the ligand (Fig 12), whereas P2 pocket is

suggested to play key role in ligand selection due to two prominent residues -His685 & His686 (Fig15).

Given the PDB co-ordinates as input **CASTp** [Liang J et al. 1998, Binkowski. T.A et al., 2003, Dundas J et al., 2006], an online tool; provides a comprehensive and detailed quantitative characterization of interior voids and surface pockets of proteins 3-D structure Details about the putative binding site as obtained from CASTp are shown (Fig 22 & table 9).

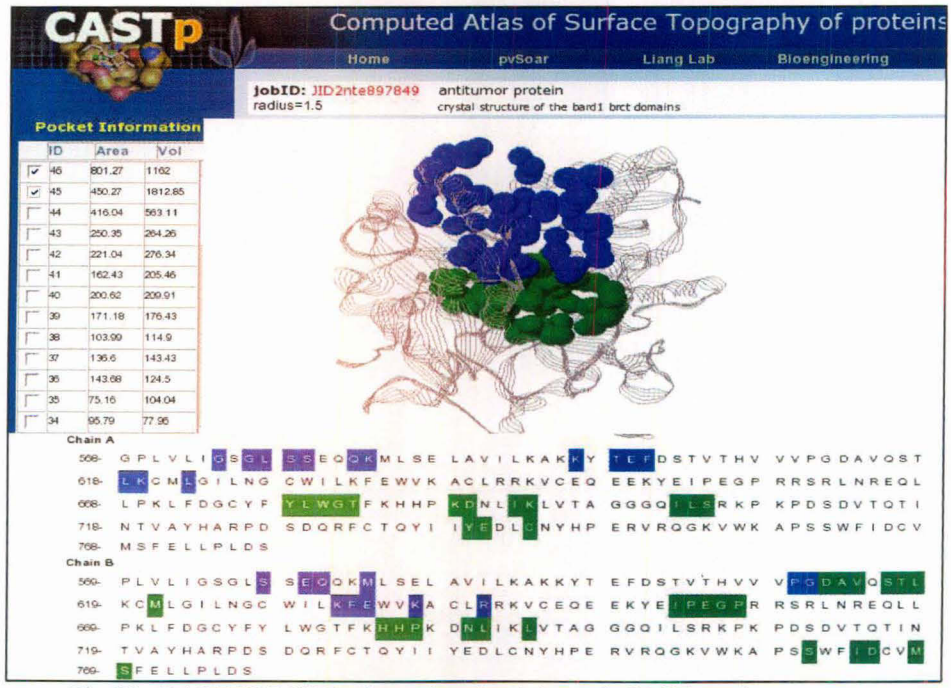


Fig 22: BARD1BRCT Pockets visualized through CASTp web server

Table 9: Active site detection by CASTp web server & literature.

Web server	Pocket ID	Rank	Common Active Residue	Area	Volume	Literature
CASTp	46	1 (green)	Ser616,Met621, His685,His686,Ile764	801.27	1162	Pocket P2
	45	2 (blue)	Gly576,Lys619	450.27	1812.86	Pocket P1

6.2 Preparation of protein

Co-ordinates of two proteins crystal structures were downloaded from RCSB Protein Data Bank website (http://www.rcsb.org). These are 1. Lysozyme co-crystallized with Isobutyl Benzene (pdb id: 184L.pdb) and 2. BARD1- BRCT domain (pdb id: 2nte.pdb). Preparation of Proteins i.e. addition of charges, addition of Hydrogens, optimization and minimization of energy were done using respective tools provided with the docking software.

6.3 Database for Virtual Screening

Version 8 of ZINC [Irwin J.J et al., 2005]; a free Database of commercially available compounds for virtual Screening (http://zinc.docking.org) was used in this study. Currently, ZINC has ~8 million compounds (the size of this library continues to grow) each with 3D structure, using catalogs of compounds from vendors. Each molecule in the library contains vendor and purchasing information and is ready for docking using a number of popular docking programs. The molecules have been assigned biologically relevant protonation states and are annotated with properties such as molecular weight, calculated LogP, and number of rotatable bonds. This database is available for free download in several common file formats including SMILES, mol2, 3D SDF, and DOCK flexibase format. A subset of compound in file 3_p0.100.sdf was downloaded. It contains 19924 molecules each with unique ZINC ID. These compounds were further filtered using tools from Open Eye Scientific software.

6.4 FILTER in virtual screening

Tools provided by Open Eye version 2.2.5 was used in this study. FILTER is an application for filtering large compound sets based on user-supplied or default rules to remove compounds that have undesirable properties (table 10). A subset of a commercially available compound 3_p0.100.sdf containing 19924 molecules from ZINC compound database was filtered applying Lipinski filter criteria. Finally 8132 drug-like molecules passed the filter which was docked using four docking tools and the chosen proteins.

Table 10: Lipinski Filter criteria [Lipinski C.A et al., 2001]

Lipinski Rule of Five	
Molecular weight (MWT)	< 500
Partition coefficient (xLogP)	< 5
No. of H bond donor (N_h_donors)	< 5
No. of H bond acceptor (n_h_acceptors)	< 5x2

6.5 Virtual Screening using Docking tools:

6.5.1 Docking Algorithms

GLIDE. Glide (Grid-Based Ligand Docking with Energetic) [Freisner R A et al., 2004, Halgren T A et al., 2004] version 8.5 was used in this study (www.schrodinger.com). Glide approximates a complete systematic (incremental) search of conformational, orientational and positional space of the docked ligand using hierarchical filters (Fig 23)

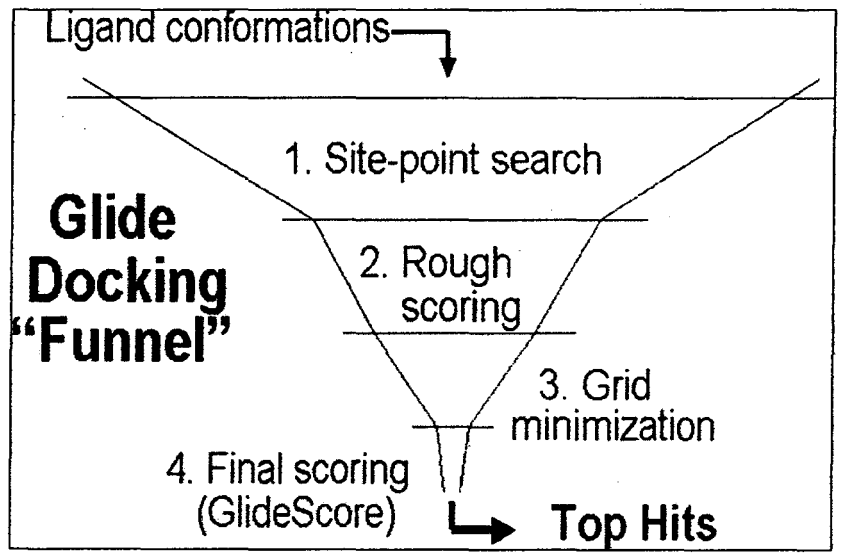


Fig 23: Glide docking "funnel", showing the Glide docking hierarchy [Freisner R.A et al., 2004].

The shape and properties of the receptor is represented on a grid by several different sets of fields that provide progressively more accurate scoring of the ligand pose. The fields are computed prior to docking. The binding site is defined by a rectangular box

confining the translations of the mass center of the ligand. A set of initial ligand conformations is generated through an exhaustive search of the torsional minima, and the conformers are clustered in a combinatorial fashion. Each cluster, characterized by a common conformation of the "core" and an exhaustive set of "rotamer group" conformations, is docked as a single object in the first stage [Freisner 2004]. The search begins with a rough positioning and scoring phase that significantly narrows the search space and reduces the number of poses to be further considered to a few hundred. In the following stage, the selected poses are minimized on precomputed OPLS-AA van der Waals and electrostatic grids for the receptor. In the final stage, the very best candidates are further refined via a Monte Carlo sampling of pose conformation which in some cases is crucial for obtaining an accurate docked pose as nearby torsional minima are examined, and the orientation of peripheral groups of the ligand is refined. The minimized poses are then rescored using the GlideScore function. The choice of the best pose is made using a model energy score (Emodel) that combines the energy grid score, GlideScore, and the internal strain of the ligand. [Freisner R. A et al., 2004].

DOCK. DOCK [Kuntz et al., 1982, Schiochet BK et al., 1992, Meng E.C et al., 1993, Ewing T.J.A et al., 1997, 2001] version 6.1 was used in this study (www.dock.compbio.ucsf.edu). DOCK, the pioneering docking algorithm of Kuntz et al., treats the geometric (hard sphere) interactions of two rigid bodies, where one body (the receptor) contains pocket or grooves that contain binding sites for the second body (the ligand), fixing the six degree of freedom (three translational and three rotational). Hence the docking problem is reduced to matching the ligand critical points to the negative image of the receptor, based on distances between the points. A minimization of the ligand poses is performed allowing DOCK to refine the ligand position in the binding pocket. For the target, the Connolly molecular surface was calculated using a probe radius of 1.4 Å. Inside the binding pocket, spheres are generated with the DOCK program SPHGEN. SPHGEN outputs the spheres in clusters that overlap each other. The sub-clustering algorithm segregates groups of spheres based on their radii. Here, a large sphere (shaded) is removed from the sphere set, removing the link connecting one part of the site to the other and segregating the spheres into two sub-clusters (Fig 24).

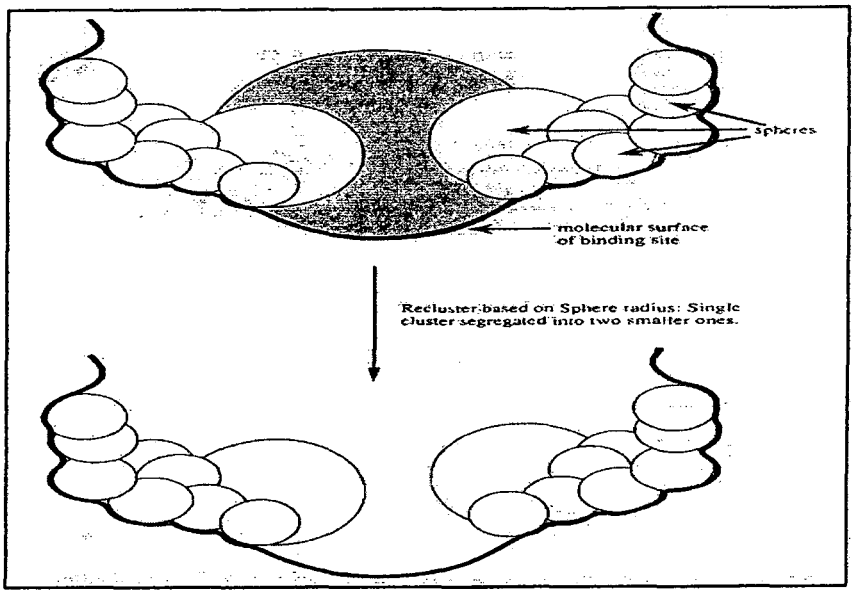


Fig 24: Sphere sub-clustering [Schiochet BK et al., 1992]

GOLD. GOLD (Genetic Optimization for Ligand Docking) [Jones G et al., 1997, Verdonk M.L et al., 2003, 2005], Version 4.0 was evaluated in the present study **(www.ccdc.cam.ac.uk)**. The GOLD program uses a genetic algorithm (GA) (Fig 25) to explore the full range of ligand conformational flexibility and the rotational flexibility of selected receptor hydrogen. The mechanism for ligand placement is based on fitting points. The program adds fitting points to hydrogen-bonding groups on the protein and maps acceptor points in the ligand on donor points in the protein and vice versa.

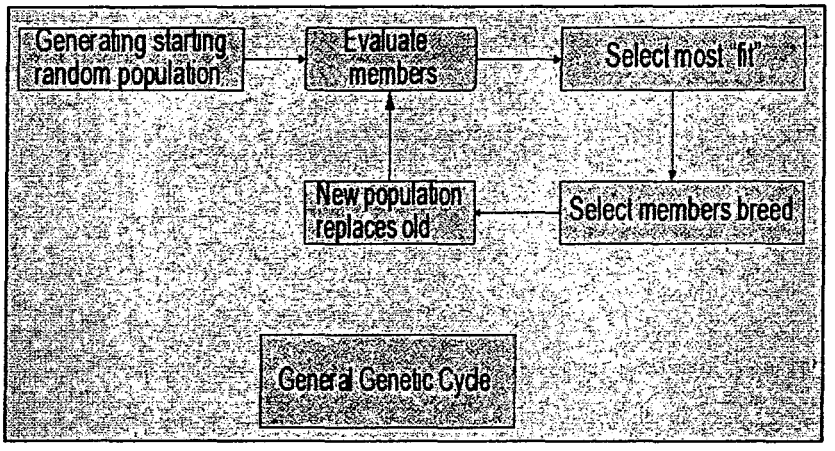


Fig 25: Sequence of events in General Genetic Cycle algorithm [Parill A L, 1996]

The docking poses are ranked based on a molecular mechanics-like scoring function. There are two different built-in scoring functions in the GOLD programs GoldScore and ChemScore. The ChemScore function implemented in GOLD is an optimized version of the original ChemScore function developed by Eldridge et al., 1997. The Simplex algorithm (local optimization) is used to relax each docking in the alternative scoring function.

FRED. FRED [McGann MR et al., 2003] docking calculations were performed using version 2.2.4 (www.openeye.com). For efficient handling of large compound databases, FRED distributes jobs via Parallel Virtual Machine (PVM) [Geist Al et al., 1994] over multiple processors. The first stage in docking is a shape fitting process, which takes a set of ligand conformers as input and tests them against a "bump map" (a Boolean grid with true values where ligand atoms can potentially be placed). Orientations that clash with the protein or are distant from the active site are rejected. The crude docking solutions are further tested against a pharmacophore feature if specified, and any poses that do not satisfy the pharmacophore are rejected. Poses surviving the shape fitting routine can then be passed through up to three scoring function filters in the screening process. Various options are available for optimization with respect to the built-in scoring functions: optimization of hydroxyl group rotamers, rigid body optimization, torsion optimization, and reduction of the number of poses that are passed on to the next scoring function. Available scoring functions in FRED are ChemScore, PLP, ScreenScore, and Gaussian shape fitting. The latter is a proprietary function of Open Eye. Qualitatively, the Gaussian scoring function has favorable values when the ligand and protein have a high surface contact and little volume overlap.

6.5.2 Scoring Functions

GlideScore. The principal scoring function used in Glide [Freisner R A et al., 2004, Halgren T A et al., 2004] is called GlideScore, a modified version of Empirical scoring function ChemScore [Eldridge et al., 1997] with slightly different weighting factors for each term and an additional terms accounting for solvation and repulsive interactions. ChemScore function of Eldridge et al., can be written as:

$$\Delta G_{bind} = C_0 + C_{lipo}\sum f(r_{lr}) + C_{hbond}\sum g(\Delta r)\, h(\Delta\alpha) + C_{metal}\sum f(r_{lm}) + C_{rotb}H_{rotb} \quad (1)$$

The summation in the second term extends over all ligand-atom/receptor-atom pairs defined by ChemScore as lipophilic, while that in the third term extends over all ligand-receptor hydrogen-bonding interactions; the fourth over the metal-ligand interactions and fifth term denotes rotational bonds. In eq1: f, *g* and *h* are functions that give a full score (1.00) for distances or angles that lie within nominal limits and a partial score (1.00-0.00) for distances or angles that lie outside those limits but inside larger threshold values. For example, *g* (Δr) is 1.00 if the H...X hydrogen bond distance is within 0.25 Å of a nominal value of 1.85 Å but tails off to zero in a linear fashion if the distance lies between 2.10 and 2.50 Å. Similarly, *h* ($\Delta\alpha$) is 1.00 if the Z-H...X angle is within 30° of 180° and decreases to zero between 150° and 120°.

Glide employs two forms of GlideScore: (i) Glide-Score SP, used by Standard-Precision Glide; (ii) GlideScore XP [Friesner R. A et al., 2006], used by Extra-Precision Glide. These functions use similar terms but are formulated with different objectives. Specifically, GlideScore SP is a softer, more forgiving function that is adept at identifying ligands that have a reasonable propensity to bind, even in cases in which the Glide pose has significant imperfections. This version tries to minimize false negatives and is appropriate for many database screening applications. In contrast, GlideScore XP [Friesner R. A et al., 2006], is a harder function that exacts severe penalties for poses that violate established physical chemistry principles such as that charged and strongly polar groups be adequately exposed to solvent. This version of GlideScore is more adept at minimizing false positives and can be especially useful in lead optimization or other studies in which only a limited number of compounds will be considered experimentally and each computationally identified compound needs to be as high in quality as possible. GlideScore (eq2) modifies and extends the ChemScore (eq1)

$$\Delta G_{bind} = C_{lipo\text{-}lipo}\sum f(r_{lr}) + C_{hbond\text{-}neut\text{-}neut}\sum g(\Delta r)\, h(\Delta\alpha) + C_{hbond\text{-}neut\text{-}charged}\sum g(\Delta r)\, h(\Delta\alpha) + C_{hbond\text{-}charged\text{-}charged}\, g(\Delta r)\, h(\Delta\alpha) + C_{max\text{-}metal\text{-}ion}\sum f(r_{lm}) + C_{rotb}H_{rotb} + C_{polar\text{-}phob}V_{polar\text{-}phob} + C_{coul}E_{coul} + C_{vdW}E_{vdW} + \text{solvation terms} \quad (2)$$

The lipophilic-lipophilic term is defined as in ChemScore. The hydrogen-bonding term also uses the ChemScore form but is separated into differently weighted components that depend on whether the donor and acceptor are both neutral, one is neutral and the other is charged, or both are charged. The metal-ligand interaction term (the fifth term in eq 2) uses the same functional form as is employed in ChemScore but varies in three principal ways. First, this term considers only interactions with anionic acceptor atoms (such as either of the two oxygen of a carboxylate group). Second, Glide counts just the single best interaction when two or more metal ligations are found. Third, the net charge on the metal ion in the unligated apo protein (generally via examination of the directly coordinated protein side chains) is assessed. The seventh term, from Schrodinger's active site mapping facility, rewards instances in which a polar but non-hydrogen-bonding atom (as classified by ChemScore) is found in a hydrophobic region (ChemScore does not penalize mismatches between lipophilic and hydrophilic groups).

The second major component is the incorporation of contributions from the Coulomb and vdW interaction energies between the ligand and the receptor.

The third major component is the introduction of a solvation model. Previous versions of GlideScore did not properly take into account the severe restrictions on possible ligand poses that arise from the requirement that charged and polar groups of both the ligand and protein are adequately solvated. Charged groups, in particular, require very careful assessment of their access to solvent. In addition, water molecules may be trapped in hydrophobic pockets by the ligand, also an unfavorable situation. To include solvation effects, Glide docks explicit waters into the binding site for each energetically competitive ligand pose and employ empirical scoring

terms that measure the exposure of various groups to the explicit waters. This "water-scoring" technology has been made efficient by the use of grid-based algorithms. Using explicit waters, as opposed to a continuum solvation model, have significant advantages. In the highly constrained environment of a protein active site containing a bound ligand, the location and environment of individual water molecules become important. Current continuum solvation models have difficulty capturing these details, but explicit-water approach has allowed developing consistently reliable descriptors for rejecting a high fraction of the false positives that appear in any empirical docking calculation. This analysis also produced trial values for the coefficients of the various penalty terms which are described in later version of glide [Freisner, 2006].

GridScore. DOCK ver. 6.1 docked the compounds allowing for rigid docking using the grid-based energy scoring option for minimization after initial placement in the site. Clusters were examined for the target, and the cluster covering the binding site (or active residues) was chosen by selecting those spheres within 10 Å of the reference ligand. The box for the scoring grid was defined such that all spheres were enclosed with an extra 5.0 Å added in each dimension. Scoring grids for contact and energy scores were calculated with a grid spacing of 0.3 Å. The bump check was set such that compounds with atoms closer than half the sum of the van der Waals radii of the respective atoms were rejected. The radii used were those in the vdw_AMBER_parm99.defn set.

GoldScore. The fitness functions offered by GOLD are: GoldScore, ChemScore, ASP (Astex Statistical Potential) and User Defined Score. In this study GoldScore fitness function was chosen. The GOLD fitness function is made up of four components:

- Protein-ligand hydrogen bond energy (external H-bond)
- Protein-ligand van der Waals (vdw) energy (external vdw)
- Ligand internal vdw energy (internal vdw)
- Ligand torsional strain energy (internal torsion)
- Optionally, a fifth component, ligand intramolecular hydrogen bond energy (internal H-bond), may be added.

Gaussian Scoring Function. Scoring functions implemented in FRED are Gaussian shape scoring, ChemScore, PLP and ScreenScore. The Gaussian shape function [McGann MR et al., 2003] describes the shape of individual atoms as spherical Gaussian functions and returns favorable score values for a large surface contact between ligand and receptor and low volume overlap. ScreenScore was derived through a combination of PLP and FlexX terms. The ScreenScore implementation in FRED does not include an angular term for metal contacts, and features an additional clash term that penalizes heavy atom clashes with less than 0.5 Å overlap by 1 kJ/mol, and more severe clashes by 10 kJ/mol. The overlap is calculated as the difference between the atom distances and the sum of the atom van der Waals radii.

6.5.3 Docking Protocols and parameters

For simplification, in all algorithms studied here, the receptor is treated as rigid body and commonly used scoring functions are considered for evaluation. This consideration was done to allow performing docking efficiently especially for database screening.

Glide. The subset of tools in Glide used for Docking is 1.Protein preparation wizard 2.LigPrep (for ligand preparation) 3. Receptor Grid Generation 4.Glide docking. Parameters used for the study are: Force Field: OPLS-AA (Optimized Potential for Liquid Simulation-All Atom), Ligand maximum atoms 200, Maximum rotatable bonds 35, VdW radii of ligand atoms scaled by 0.800000, Charge cutoffs for polarity 0.150000, Receptor setup: (nsites, nx, ny, nz, bsize) = (216, 22, 22, 22, 1.0000). Penalizing non-planar amide torsions- Number of rotatable bonds 7, Core rotatable bonds, max cores 6, 150. GlideScore version XP 5.0, Buried polar penalty 0.000, Coulomb vdW cutoff 0.000, H bond cutoff 0.000, Metal-ligand cutoff 10.000, For Glide Extra Precision mode, maxref = 800

DOCK. Docking protocol is divided into four parts 1. Structure Preparation 2. Sphere generation and selection 3. Grid Generation 4. Docking. Each is mentioned in detail:

1. Structure preparation:

For protein structure:

Deletion of ligand from receptor crystal structure

Addition of hydrogen

Add charge to each atom of receptor molecule

Generation of Mol2 file from PDB file format

For ligand:

Addition of hydrogen

Add charge to ligand atoms

Above was done by using program called Babel using command

babel <option> -i<format> inputfile -o<format> output file

e.g.: to convert the PDB file to MOL2 file

babel -ipdb 184L.pdb -omol2 184L.mol2

With the option; -h add hydrogen, -d Delete hydrogen. Alternatively, addition of hydrogen and charge can be done using Chimera.

2. Sphere generation and selection

Step1: Generate the molecular surface of the receptor

For sphere generation molecular surface of the target was generated using the program called DMS. E.g.: $DMS/dms rec_noH.pdb -n -w 1.4 -v -o rec.ms

-n calculate normals for surface points

-w change probe radius

-v verbose

-o output file name

Step2. Generation of spheres surrounding the receptor:

To generate spheres from the molecular surface and the normal vectors, the program sphgen distributed as an accessory with DOCK was used. Spheres were calculated over the entire surface, producing approximately one sphere per surface point. This dense representation was then filtered to keep only the largest sphere associated with each surface atom. The filtered set was then clustered using a single linkage algorithm. Each resulting cluster represented an evagination in the target. The sphgen input file must be named INSPH, and contained following information:

r #molecular surface file

R #sphere outside of surface (R) or inside surface (L)

X #specifies subset of surface points to be used (X=all points)

0.0 #prevents generation of large spheres with close surface contacts (default=0.0)

4.0 #maximum sphere radius in Angstroms (default=4.0)

1.4 #minimum sphere radius in Angstroms (default=radius of probe)

rec.sph #clustered spheres file

To generate the spheres, the command "sphgen" in the same folder that contains the INSPH and the rec.ms files was used. Two output files were: rec.sph, which contains the spheres in clusters, and OUTSPH, which contains general information about the calculation.

Step3: Selection of a subset of spheres to represent the binding site(s)

Selected spheres within some radius of a desired location. If the active site is known then spheres can be selected within a radius of a set of atoms that describes the site. This was done using the program sphere selector, which is distributed as an accessory with DOCK. The syntax for sphere selector is:

sphere_selector <sphere_cluster_file.sph> <set_of_atoms.mol2> <radius>

3. Grid generation

The charged.mol2 and the selected_spheres.sph files were obtained from "Structure Preparation" and the "Sphere Generation and Selection".

Step 1: Creating a box around the active site

The command "showbox" was run to generate the question tree and calculate the grid box. Alternatively, answers to the questions can be listed in a text file, which can then be piped into showbox with the command "showbox < box. in."

Step 2: Generation of the Grid

To generate the grid the program grid that is distributed as an accessory to DOCK was used. Using the box generated in Step1, the program grid pre-computes the contact and electrostatic potentials for the active site at specified grid spacing. In order to run grid, a file named grid. in was generated either interactively by answer questions or in a text file. grid -i input_file [-o output_file]

4. Docking

The lig_charged.mol2 files were obtained from the "Molecule Preparation", the selected_spheres.sph file from the "Sphere Generation and Selection" and the grid.nrg and grid.bmp files from the "Grid Generation". The program dock6 distributed and installed with DOCK was required. This ligand input file containing single or multiple ligands was treated simply by including more ligands in the ligand input file.

Rigid ligand docking

Out of two docking options-rigid and flexible docking available with DOCK v6.1, rigid docking was performed keeping all the default setting. In this first case, the ligand will be kept completely rigid during the orientation step. The purpose is to explore the matching and minimization algorithms.

To actually run the docking calculation, the program dock6 was needed that is distributed with DOCK. An input file- rigid.in was generated. The command to run docking is:

dock6 -i dock.in [-o dock.out]

Parameters used for running DOCK 6.1 were as follows: DOCK requires following receptor files: a MOL2 format file containing coordinates of all atoms; a PDB file containing heavy atoms coordinates only; a PDB file containing heavy atoms excluding those of the active site. The active site atoms included those receptor atoms that were within 10 Å from the reference ligand atoms. The ligand coordinates were provided in MOL2 format. The site points for the ligand docking were identified using SPHGEN program. The energy score were employed for the orientational and conformational search. Grid maps were calculated using the program Grid, with grid spacing of 0.3 Å. An energy cutoff distance of 10 Å was employed. Electrostatic interactions were calculated with distance dependent dielectric constant. The dielectric factor was set to be 4. Proteins were represented by all atom models. Flexible bonds and anchors were automatically identified by DOCK. The conformational search was done using torsional driver. The bump check was set such that compounds with atoms closer than half the sum of the van der Waals radii of the respective atoms were rejected. The bump overlap was set to 0.75. (Amount of VDW overlap allowed. If the probe atom and the receptor heavy atom approach. #closer

than this fraction of the sum of their VDW radii, then the position is flagged as a bump; 0 = Complete overlap allowed, 1 = No overlap allowed). The radii used were those in the vdw_AMBER_parm99.defn set.

For T4 Lysozyme (pdb id. 184L.pdb)

Total charge on 184L.pdb : 9.002

Grid box information.

Box center of mass : 25.961 8.129 4.286

Box dimensions : 29.091 33.166 31.387

Number of grid points per side [x y z]: 98 112 106

Total number of grid points : 1163456

Defining reference ligand for BARD1- BRCT domain

Since the target protein BARD1- BRCT domain doesn't contain any known bound ligand, molecule showing topmost GScore by Glide SP (i.e. ZINC08928542) was taken as reference. Clusters were examined for the target, and the cluster covering the binding site (or active residues) was chosen by selecting those spheres within 10 Å of the co crystallized ligand. Following same protocol and parameters which was used for T4 Lysozyme, the information obtained for P2 pocket of BARD1- BRCT is as follows:

For BARD1- BRCT P2 pocket (pdb id. 2NTE.pdb)

Total charge on bardp2_comp.pdb : 7.000

Grid box information

Box center of mass : 17.569 8.471 16.487

Box dimensions : 34.333 29.514 32.281

Total number of grid points : 1264400

GOLD. GOLD requires the receptor and ligand co-ordinates in any of the following formats: PDB, MOL, SDF or MOL2 format. The active site origin was specified by the centre of geometry of the reference ligand in case of T4 Lysozyme and by providing atom number of active residue as in case of BARD1- BRCT domain. In the present work, the binding site was defined as a spherical region which encompasses all protein atoms within 10.0 Å of an active residue referenced from crystallographic

ligand or active site residue searched from literature and verified by CASTp, an online cavity search tool. Following default parameters were used for all calculation: Population size = 100, Selection Pressure = 1.1, No. of Islands = 5, Maximum operations = 100,000, Niche size = 2. Genetic Operators- Point crossover = 95, Migration = 10, Genetic Algorithm (GA) run = 10, Flood Fill- Radius = 10, cavity_file = .mol2, floodfill_center = cavity_from_ligand (In case of BARD1BRCT domain atom no. from putative binding pocket was provided). TERMINATION- early termination, no. of top solutions = 3, rms tolerance = 1.5. Fitness Function Settings- initial_virtual_pt_match_max=3, relative_ligand_energy = 0, start_vdw_linear_cutoff = 6, score_param_file = Default.

FRED. FRED require input format for ligand as .mol2 whereas protein as .oeb.gz. Charges were assigned to protein and ligand. Parameters used were as follows: 5Å added to the active site residues. Docking: Exhaustive Search, exhaustive scoring chemgauss3, number of poses 100. Standard Scoring Functions were: shapegauss, plp, chemgauss3, oechemscore, screenscore, zapbind. Output format is: sdf.gz, -hitlist_size 1000. Default settings were used for all the calculations.

6.6 Validation of docking results

The following tools were used for analyzing the high ranked docked molecules.

Getneares: Getneares, a tool of DOCK version5.1.1 calculates the protein atoms that are involved in making an interaction with the ligand atoms within the range of 5 Å radius. The output is written in a file containing all atoms of all residues in receptors, which has any atom within specified distance of ligand. The closest receptor residues that interact with the ligand atom, is written in .lst file.

Ligplot: Ligplot is used for plotting the interaction between receptor protein and ligand. The hydrogen bonded interaction and non hydrogen bond interaction are tabulated on output file. In order to identify the key interaction, Ligplot plots the interaction as a 2D plot.

CHAPTER 7.

RESULTS

7.1 Docking protocols

To derive optimized parameters for docking using different methods here a standard protein T4 Lysozyme, a single domain well-known active site with many 3D structures available with inhibitors was used (Fig 26).

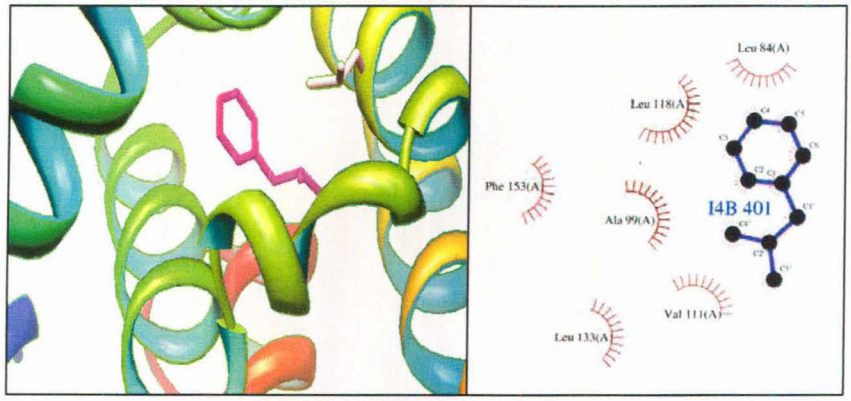


Fig 26: Crystal structure of **T4 Lysozyme** with Known bound inhibitor Isobutyl benzene and interaction of T4 Lysozyme residues with ligand [Tool: Ligplot]

T4 Lysozyme co-crystallized with inhibitor Isobutyl benzene (pdb id 184L.pdb) were redocked using four docking methods GOLD, Glide, DOCK & FRED. 3-D structures are mainly viewed through UCSF Chimera. The methods and parameters used for each docking tool are described in Material and method (chapter 6).

APPROACH

1.'Best pose' & 'Best scored pose'

i) X-ray pose (Re docking)

ii) Oriented with &/or without constraints

iii) Oriented & translated with &/or without constraints

2. Discrimination between known binders from randomly chosen molecule for vs.

To show best 'best pose' and 'best scored pose', all the HET atoms were removed from T4 Lysozyme (pdb id.184L.pdb). Co-ordinates of Isobutyl benzene were saved separately as lig.pdb. The co-ordinates of original ligand (lig.pdb) were changed in several different ways and each saved separately in .pdb and/or .mol2 format using Chimera. This changes include-ligand oriented without constrains (lig_orn1.pdb) and with constraints i.e. rotation along X-axis (ligrot1.pdb), ligand oriented and translated without constraints (lig_orntrans.pdb) and with constraints (ligrotrans.pdb) (Fig 27a, b)

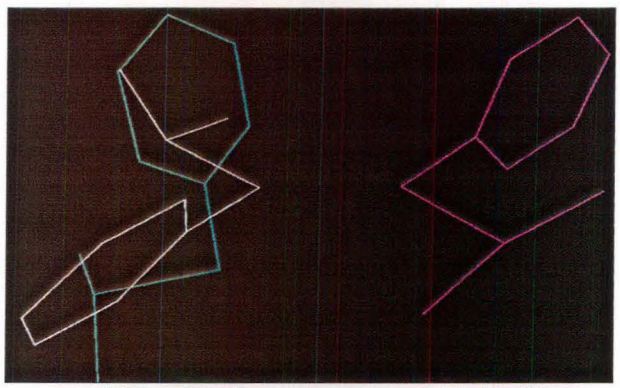


Figure 27a: X-ray pose of ligand Isobutyl Benzene (white) without constraints- rotated (cyan), rotated and translated (pink).

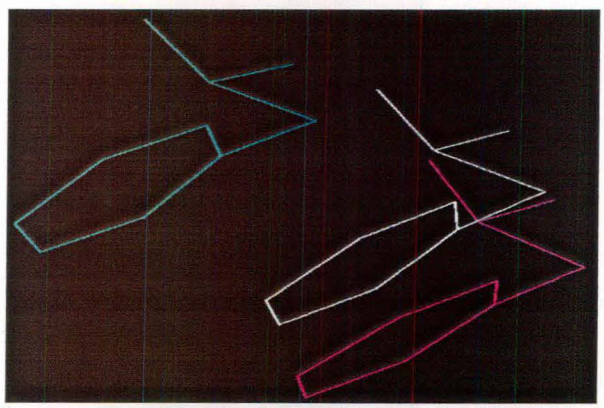


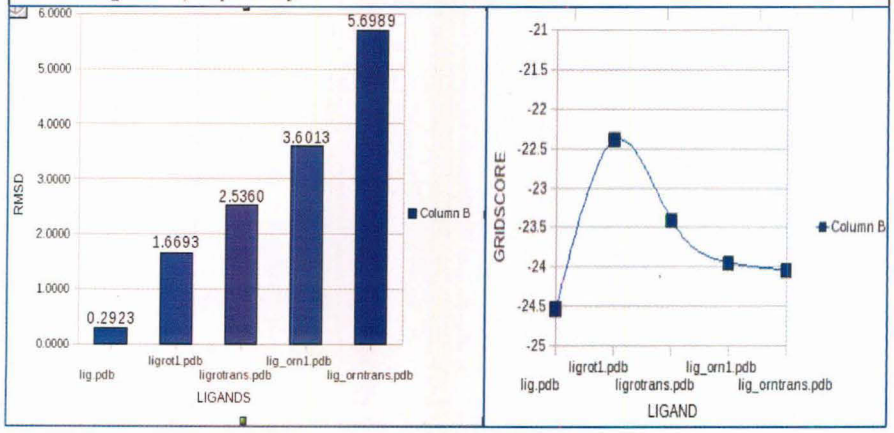
Figure 27b: Altered x-ray pose of ligand Isobutyl Benzene with constraints rotated (cyan), rotated and translated (pink).

DOCK v 6.1

Docking of Isobutyl benzene with its original and changed co-ordinates was performed in the active site of protein T4 Lysozyme. The RMSD value of isobutyl benzene with its changed co-ordinates with respect to original ligand as shown in (Fig 28a, b) explains that co-crystallized ligand pose show minimum RMSD when redocked. The changes in ligand co-ordinates were done in several ways i.e. rotated, rotated and translated with and without constrains (Fig 29a, b). RMSD with respect to co-crystallized ligand gradually increases when the ligand with its changed co-ordinates were docked in the same active site. The GridScore was also found to be minimum in case of redocking experiment. This suggests that X-ray pose is the best pose as well as the best scored pose.

		DOCK 6			
	Name of ligand	RMSD wrt lig.pdb	Grid score	Vdw	es
i)	lig.pdb	0.292257	-24.534540	-24.421938	-0.112602
ii)	lig_orn1.mol2	3.60129	-23.944429	-23.944429	0.000000
iii)	ligrot1.mol2	1.6693	-22.392656	-22.392656	0.000000
iv)	lig_orntrans.mol2	5.69885	-24.039225	-24.039225	0.000000
v)	ligrotrans.mol2	2.53597	-23.406837	-23.406837	0.000000

i) original ligand.pdb
ii) & iii) are ligand **oriented** without and with comstraints(rotation along x-axis)
iv) & v) are ligand **oriented and tanslated** without and with comstraints(rotation & translation along x-axis) respectively.



a. Ligand vs. RMSD b. Ligand vs. Grid Score

Fig 28: Docking results of different poses of ligand Isobutyl Benzene

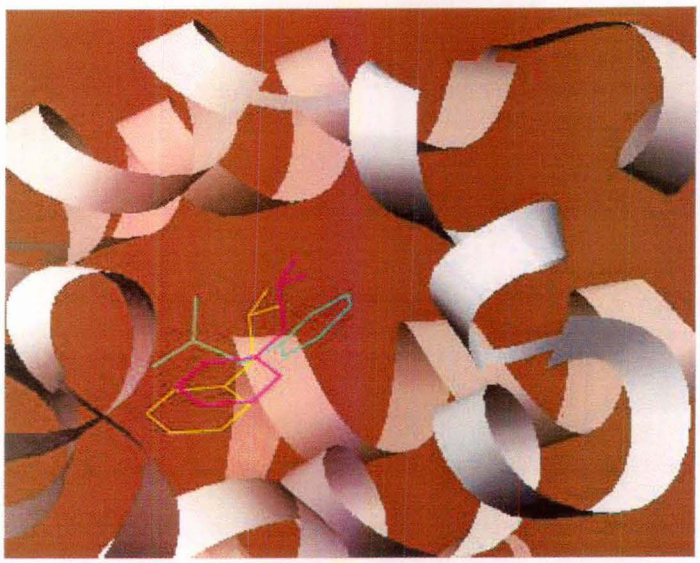


Fig 29a: Altered x-ray pose of ligand Isobutyl Benzene (yellow) i.e. rotated without (cyan) and with constraints (pink) in the active site of T4 Lysozyme (ribbon).

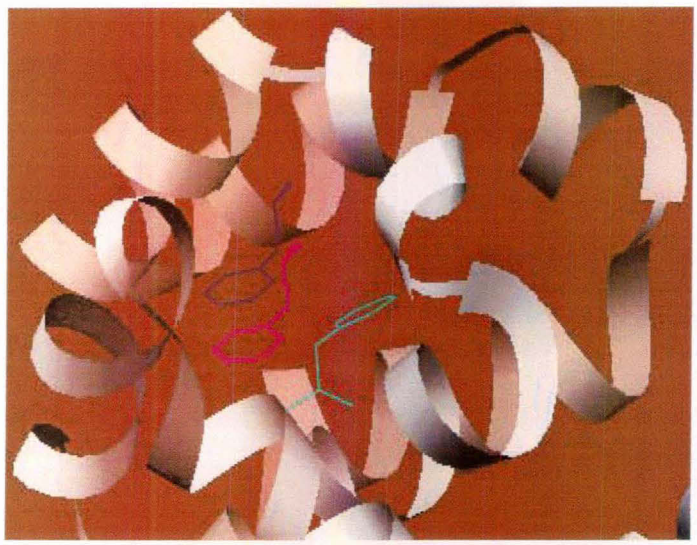


Fig 29b: Altered x-ray pose of ligand Isobutyl Benzene i.e. original (pink), rotated and translated without (cyan) and with constraint (black) in the active site of T4 Lysozyme (ribbon).

Glide docking

XP docking. Docking of Isobutyl benzene with its original and changed co-ordinates was performed in the active site of protein T4 Lysozyme [Fig 30]. The changes in ligand co-ordinates were done in several ways i.e. rotated, rotated and translated with and without constrains. Glide Extra Precision (XP) docking result (table 11) explains that GScore was not found to be minimum for X-ray pose in redocking experiment. This suggests that the pose other than X-ray pose (lig_orn1 i.e. ligand oriented without constrains) is the best scored pose.

Table 11: GScore (XP) of different poses of isobutyl benzene when docked with T4 Lysozyme.

Name of ligand	GScore
lig	-7.33
lig_orn1	-7.42
ligrot1	-7.38
lig_orntrans	-7.36
ligrotrans	-7.33

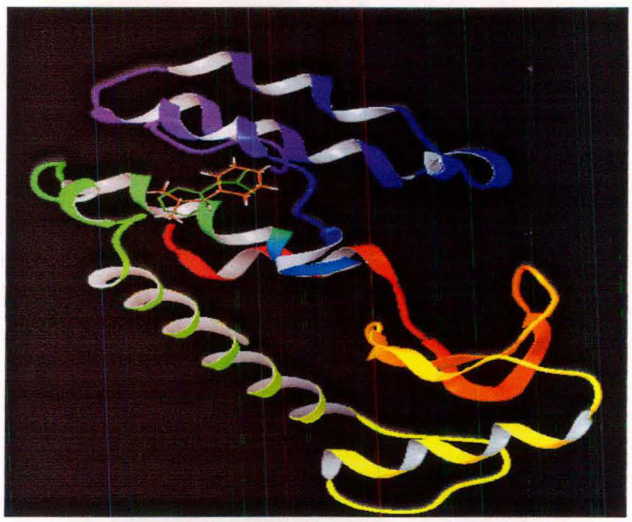


Fig 30: Glide Pose Viewer [Ribbon: T4 Lysozyme, Orange ring: original ligand (lig.pdb) Green ring: oriented & translated ligand (ligrotrans.pdb)]

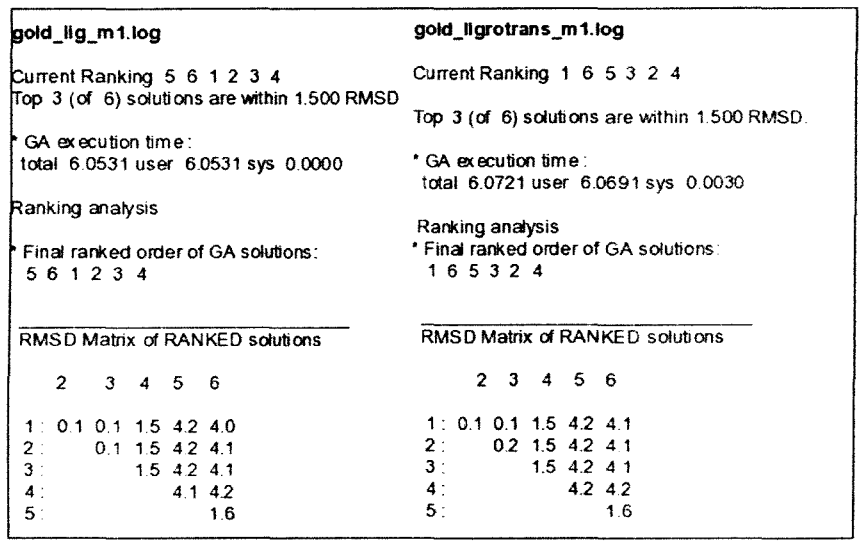
SP Docking. Glide SP docking was performed in the active site of protein T4 Lysozyme for molecules from ZINC database. GScore of top five molecules is shown in table 12. GScore of co-crystallized ligand (-7.33) is lower than the best scored ZINC database molecule, ZINC08935440 (GScore -7.06).

Table 12: Glide SP-docking for Lysozyme (GScore for co-crystallizes ligand isobutyl benzene is -7.33)

Rank	Title	GScore
1	ZINC08935440	-7.06
2	ZINC08944295	-5.86
3	ZINC08925717	-5.68
4	ZINC08938016	-5.57
5	ZINC00209314	-5.46

Gold

Docking of Isobutyl benzene with its original and changed co-ordinates was performed in the active site of protein T4 Lysozyme. The RMSD value of isobutyl benzene with its changed co-ordinates with respect to original ligand was not possible since GOLD gives RMSD of different poses of a ligand shown in (Fig 31a).



gold_lig_m1.log

Current Ranking 5 6 1 2 3 4
Top 3 (of 6) solutions are within 1.500 RMSD

* GA execution time:
total 6.0531 user 6.0531 sys 0.0000

Ranking analysis

* Final ranked order of GA solutions:
5 6 1 2 3 4

RMSD Matrix of RANKED solutions

	2	3	4	5	6
1:	0.1	0.1	1.5	4.2	4.0
2:		0.1	1.5	4.2	4.1
3:			1.5	4.2	4.1
4:				4.1	4.2
5:					1.6

gold_ligrotrans_m1.log

Current Ranking 1 6 5 3 2 4

Top 3 (of 6) solutions are within 1.500 RMSD

* GA execution time:
total 6.0721 user 6.0691 sys 0.0030

Ranking analysis
* Final ranked order of GA solutions:
1 6 5 3 2 4

RMSD Matrix of RANKED solutions

	2	3	4	5	6
1:	0.1	0.1	1.5	4.2	4.1
2:		0.2	1.5	4.2	4.1
3:			1.5	4.2	4.1
4:				4.2	4.2
5:					1.6

Fig 31a: RMSD Analysis by Gold

The changes in ligand co-ordinates were done in several ways i.e. rotated, rotated and translated with and without constrains. The GoldScore was also found to be minimum in case of redocking experiment. This suggests that X-ray pose is the best scored pose (Fig 31b).

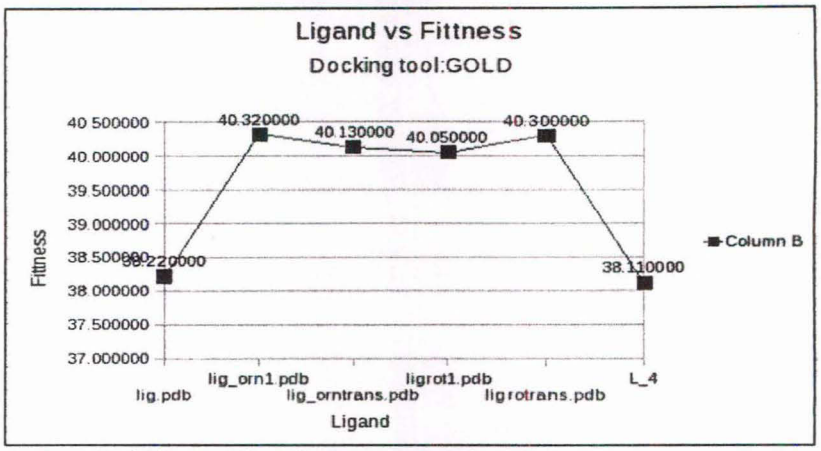
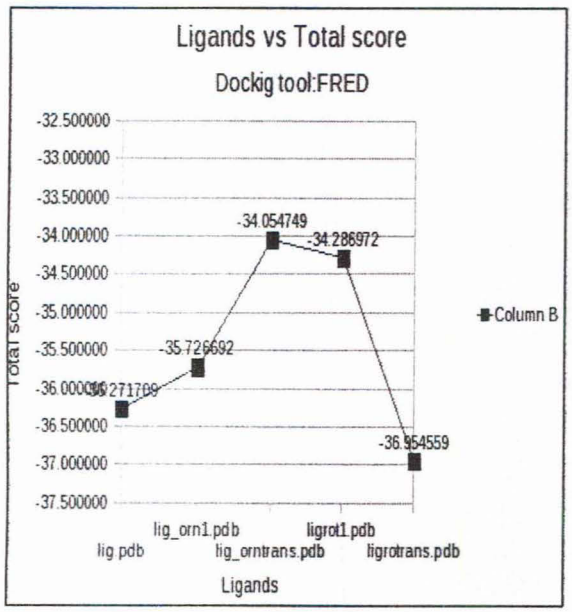


Fig 31b: Ligand vs GoldScore [lig.pdb & L_4: Original ligand]

FRED

Docking of Isobutyl benzene with its original and changed co-ordinates was performed in the active site of protein T4 Lysozyme. The changes in ligand co-ordinates were done in several ways i.e. rotated, rotated and translated with and without constrains. Total Score was not found to be minimum for X-ray pose in redocking experiment. This suggests that the pose other than X-ray pose (ligand oriented and translated with constraints (ligrotrans.pdb) is the best scored pose [Fig 32]



Ligand Name	Total score
lig.pdb	-36.271709
lig_orn1.pdb	-35.726692
lig_orntrans.pdb	-34.054749
ligrot1.pdb	-34.286972
ligrotrans.pdb	-36.954559

Fig 32: Ligand vs. total Score

7.2 BARD1BRCT domain

BARD1BRCT domain has two putative binding pockets P1 (red label) and P2 (yellow label) (Fig 33). Detailed docking studies were performed in P2 pocket of BARD1BRCT domain using all four software DOCK, Glide, GOLD and FRED while only Glide docking were performed for P1 pocket of BARD1- BRCT domain. Results were compared based on different docking algorithms and scoring functions used (Tables 13- 23, Fig 33-37).

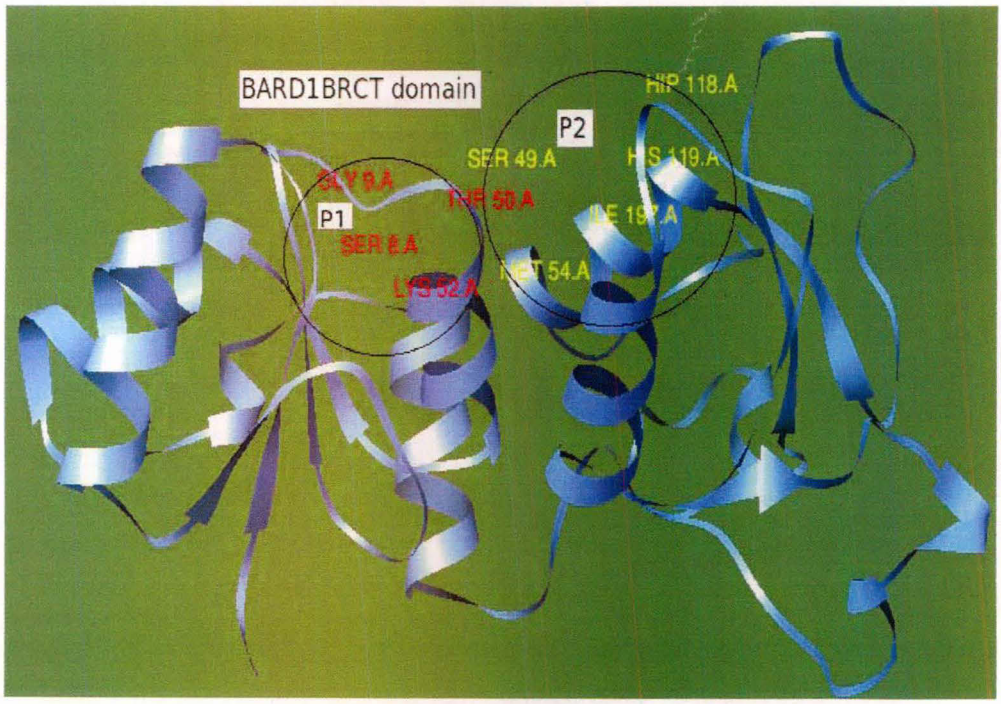


Fig 33: Binding pockets P1 (red label) and P2 (yellow label) of BARD1- BRCT domain (ribbon) are shown with the major residues are colored*.

*BARD1 BRCT domain residue 568-777 in crystal structure corresponds to residues 1-210 when visualized through Chimera (e.g. Gly 576 → Gly 9).

Table 13: List* of top 50 Score (Dock 6.1, Grid Score)

ZINC ID	Rank (Dock)	Grid Score
ZINC08933378	1	-41.64
ZINC08946869	2	-40.01
ZINC08933594	3	-39.64
ZINC08945436	4	-39.59
ZINC08939674	5	-39.34
ZINC08922692	6	-39.06
ZINC08949927	7	-38.98
ZINC08965231	8	-38.66
ZINC08922739	9	-38.64
ZINC08923686	10	-38.61
ZINC08865035	11	-38.57
ZINC08951543	12	-38.52
ZINC01834913	13	-38.49
ZINC08923491	14	-38.35
ZINC08938286	15	-38.33
ZINC08920411	16	-38.16
ZINC08958378	17	-38.07
ZINC08937677	18	-38.02
ZINC08927590	19	-38
ZINC08938057	20	-37.89
ZINC08747052	21	-37.88
ZINC08866407	22	-37.86
ZINC08865319	23	-37.81
ZINC08943410	24	-37.79
ZINC08947326	25	-37.78
ZINC08937694	26	-37.71
ZINC08796780	27	-37.58
ZINC08725793	28	-37.58
ZINC08933440	29	-37.46
ZINC08920432	30	-37.39
ZINC08946681	31	-37.38
ZINC08865915	32	-37.36
ZINC06177351	33	-37.33
ZINC08942695	34	-37.29
ZINC08938011	35	-37.26
ZINC08931864	36	-37.23
ZINC08747046	37	-37.23
ZINC08920767	38	-37.18
ZINC08937182	39	-37.15
ZINC08928019	40	-37.14
ZINC08865324	41	-37.14
ZINC08931885	42	-37.1
ZINC08865506	43	-37.1
ZINC08920375	44	-37.05
ZINC08945501	45	-37.04
ZINC08808539	46	-37.03
ZINC08933946	47	-37.03
ZINC08945579	48	-36.98
ZINC08865509	49	-36.94
ZINC08944881	50	-36.94

* Out of top 2000 molecules selected for comparative study with other tools.

Table 14: List of top 50 Fitness Score (GOLD 4.0, SF: Gold Score)

ZINC ID	Rank (Gold)	Fitness Score
ZINC03542962	1	63.7
ZINC08725805	2	63.28
ZINC08954242	3	62.42
ZINC08938597	4	62.29
ZINC08966129	5	62.23
ZINC08924955	6	61.66
ZINC08944091	7	61.55
ZINC08932815	8	61.51
ZINC08954293	9	61.24
ZINC08944717	10	61.08
ZINC08955386	11	61.03
ZINC08955886	12	60.98
ZINC08965256	13	60.69
ZINC08955899	14	60.55
ZINC08965614	15	60.44
ZINC08938480	16	60.11
ZINC08932678	17	59.95
ZINC08841054	18	59.89
ZINC08933814	19	59.44
ZINC08936423	20	59.42
ZINC08954997	21	59.37
ZINC08940029	22	59.36
ZINC08932357	23	59.31
ZINC08932686	24	59.3
ZINC08924969	25	59.26
ZINC08955206	26	59.11
ZINC08954450	27	59.08
ZINC08941015	28	59
ZINC08965417	29	58.88
ZINC08746329	30	58.77
ZINC08924546	31	58.66
ZINC08955994	32	58.59
ZINC08956008	33	58.59
ZINC08956006	34	58.45
ZINC08725981	35	58.23
ZINC08943410	36	58.23
ZINC08954297	37	58.22
ZINC08955387	38	58.13
ZINC08932353	39	58.1
ZINC08938286	40	58.08
ZINC08965711	41	58.03
ZINC08955649	42	58.01
ZINC08940024	43	58.01
ZINC08937287	44	57.98
ZINC08966265	45	57.9
ZINC08956023	46	57.9
ZINC08955108	47	57.88
ZINC08955012	48	57.85
ZINC08940850	49	57.79
ZINC08933577	50	57.73

FRED docking

The structure visualization by docking tool Fred through Graphical User interface (GUI) is run through command fred_receptor. Here, shown is active site of BARD1-BRCT domain which was used for database docking (Fig 34).

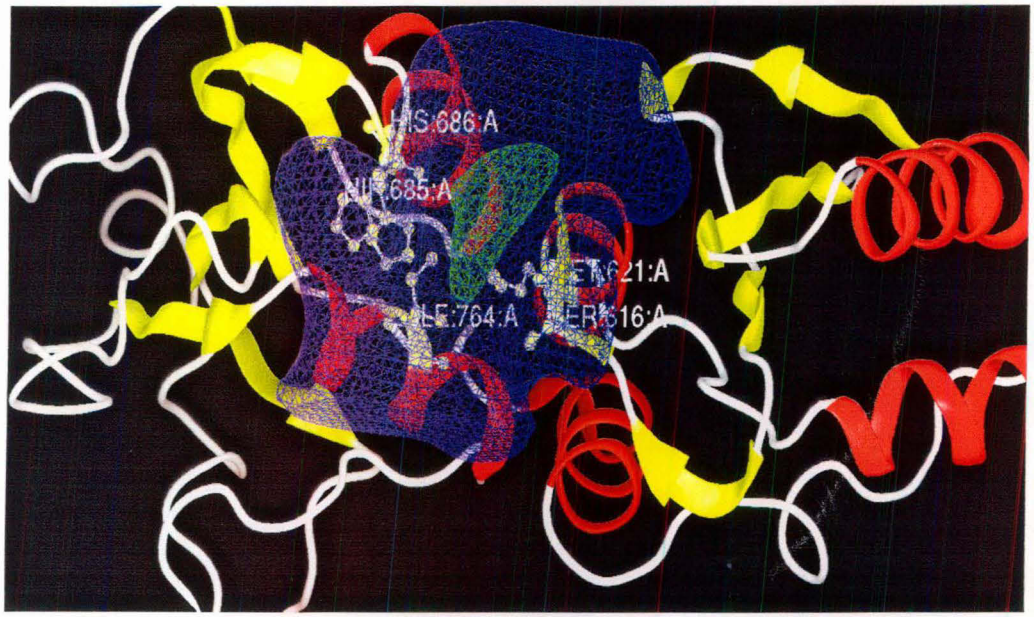


Fig 34: Active site of BARD1BRCT domain p2 pocket as visualized in Fred GUI.

Table15: List of top 50 Score (Fred: Chemgauss3)

ZINC ID	Rank (Fred)	Chemgauss3
ZINC08798570	1	-30.44
ZINC08780039	2	-29.76
ZINC08798283	3	-29.54
ZINC08945830	4	-28.5
ZINC08939945	5	-25.22
ZINC08940793	6	-24.78
ZINC08940481	7	-24.43
ZINC06813474	8	-24.21
ZINC08940164	9	-24.11
ZINC08919593	10	-23.93
ZINC08925717	11	-23.8
ZINC08746287	12	-23.76
ZINC08746898	13	-23.69
ZINC08944605	14	-23.65
ZINC08939780	15	-23.64
ZINC08924362	16	-23.33
ZINC00494897	17	-23.31
ZINC08798180	18	-23.2
ZINC08925780	19	-23.04
ZINC08918937	20	-22.85
ZINC08936016	21	-22.59
ZINC08939205	22	-22.56
ZINC08780028	23	-22.44
ZINC08747131	24	-22.37
ZINC08747108	25	-22.35
ZINC08799099	26	-22.28
ZINC08935603	27	-22.26
ZINC08966166	28	-22.24
ZINC08944180	29	-22.18
ZINC08964422	30	-22.09
ZINC08944814	31	-22.07
ZINC08919069	32	-21.98
ZINC08746855	33	-21.89
ZINC08924398	34	-21.85
ZINC08943095	35	-21.79
ZINC08942921	36	-21.76
ZINC08798278	37	-21.71
ZINC08780041	38	-21.61
ZINC08940584	39	-21.55
ZINC08778947	40	-21.53
ZINC08935355	41	-21.09
ZINC08798294	42	-21
ZINC08919823	43	-20.95
ZINC04256153	44	-20.92
ZINC08924511	45	-20.91
ZINC08919079	46	-20.91
ZINC08942407	47	-20.61
ZINC08746809	48	-20.53
ZINC08945920	49	-20.52
ZINC08919072	50	-20.52

Table16: List of top 50 GScore (Glide XP), comparison with GoldScore, GridScore, Chemgauss3 & their ranks [R.no]

ZINC ID	Gscore	R no	Gold Score	R no	Grid Score(Dock)	R no	Chemgauss3(Fred)	R no
ZINC08939981	-6.24	1	39.56	5641				
ZINC08725662	-6.18	2	42.76	4250				
ZINC08746714	-6.08	3	50.2	997				
ZINC08931589	-5.64	4	41.63	4788				
ZINC08746724	-5.57	5	42.27	4505				
ZINC06622854	-5.5	6	54.66	168				
ZINC08725914	-5.48	7	42.65	4318				
ZINC08945649	-5.37	8	41.82	4710				
ZINC08746434	-5.33	9	43.65	3856	-32.59	1359		
ZINC08798104	-5.33	10	46.38	2530				
ZINC06623086	-5.27	11	41.64	4778				
ZINC08923736	-5.25	12	39.06	5816				
ZINC06623665	-5.23	13	37.29	6332	-11.72	346		
ZINC00626488	-5.22	14	40.37	5351				
ZINC08746714	-5.17	15	50.2	997				
ZINC08746862	-5.17	16	42.86	4208				
ZINC08799104	-5.06	17	48.74	1480				
ZINC08746968	-5.03	18	42.49	4388				
ZINC08799099	-5.03	19	41.8	4714	-35.07	250	-22.28	26
ZINC08747178	-5.03	20	43.52	3911				
ZINC08799001	-5.01	21	42.34	4465				
ZINC08725662	-5.01	22	42.76	4250				
ZINC08799104	-5	23	48.74	1480				
ZINC08799099	-4.98	24	41.8	4714	-35.07	250	-22.28	26
ZINC08945923	-4.96	25	41.12	5032	-6.2	497		
ZINC08725662	-4.93	26	42.76	4250				
ZINC03105140	-4.93	27	51.01	774				
ZINC08798793	-4.92	28	50.36	946				
ZINC08725914	-4.92	29	42.65	4318				
ZINC08798104	-4.9	30	46.38	2530				
ZINC08799099	-4.9	31	41.8	4714	-35.07	250	-22.28	26
ZINC08798015	-4.9	32	44.48	3429	-33.02	1083		
ZINC03105143	-4.87	33	49.54	1193				
ZINC08938257	-4.86	34	44.77	3288				
ZINC02192105	-4.85	35	44.37	3491				
ZINC08798109	-4.84	36	48.54	1553				
ZINC08778698	-4.84	37	52.58	435	-33.29	915		
ZINC00793994	-4.8	38	52.02	546				
ZINC08798418	-4.79	39	40.67	5231				
ZINC00626488	-4.77	40	40.37	5351				
ZINC08798776	-4.76	41	47.14	2163				
ZINC08924594	-4.75	42	50.84	823	-2.08	710		
ZINC06623665	-4.74	43	37.29	6332	-11.72	346		
ZINC08798750	-4.73	44	42.16	4539				
ZINC08918941	-4.73	45	28.12	7691	-15.44	317		
ZINC08747177	-4.71	46	42.3	4482				
ZINC08937692	-4.71	47	48.2	1696				
ZINC08864735	-4.7	48	40.17	5421				
ZINC08939609	-4.7	49	37.27	6340				
ZINC03490351	-4.7	50	48.24	1672				

Table 17: List of top 50 GScore (GlideXP) for BARD1BRCT P1 pocket

ZINCID	Gscore
ZINC08937526	-4.08
ZINC08937478	-4.04
ZINC08935841	-3.83
ZINC08746717	-3.61
ZINC08746720	-3.54
ZINC08928323	-3.49
ZINC08865783	-3.31
ZINC08932447	-3.29
ZINC08937631	-3.18
ZINC08937799	-3.13
ZINC08933955	-3.09
ZINC08933859	-3.03
ZINC08937914	-3
ZINC08965788	-4.02
ZINC00316306	-4.01
ZINC06623665	-3.77
ZINC08955479	-3.41
ZINC08959384	-3.21
ZINC00305594	-3.2
ZINC08866406	-3.13
ZINC08964427	-3.13
ZINC00793994	-3.02
ZINC08923736	-3.79
ZINC08921432	-3.69
ZINC02561446	-3.67
ZINC06813474	-3.59
ZINC08746712	-3.57
ZINC02561432	-3.39
ZINC08746293	-3.38
ZINC08920573	-3.34
ZINC08927408	-3.29
ZINC08925246	-3.25
ZINC08808517	-3.21
ZINC08746269	-3.19
ZINC08921892	-3.11
ZINC08924379	-3.1
ZINC08926748	-3.08
ZINC08924410	-3.07
ZINC08924895	-3.03
ZINC08922295	-3.03
ZINC08921818	-3.01
ZINC02561430	-3.01
ZINC08919353	-3
ZINC08939630	-3.66
ZINC08944988	-3.44
ZINC08940733	-3.43
ZINC08944280	-3.22
ZINC08945649	-3.21
ZINC08944794	-3.19
ZINC08945966	-3.1

The Color code for fig 35 onwards: Red: Oxygen; Blue: Nitrogen; White: Hydrogen; Dash line: H bond interaction with distance in Å; Ribbon: Protein; BARD1- BRCT domain is shown.

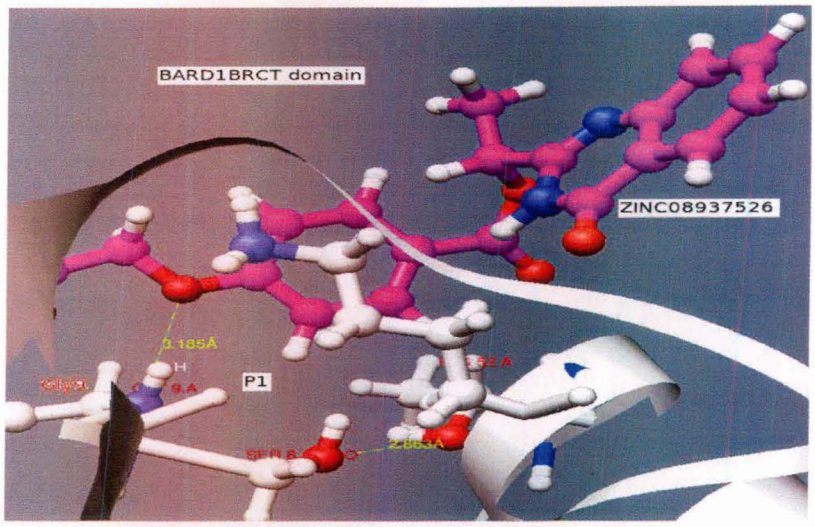


Fig 35: Molecule with ZINC ID- 08937526 ranked 1 for P1 pocket of BARD1BRCT domain by Glide XP

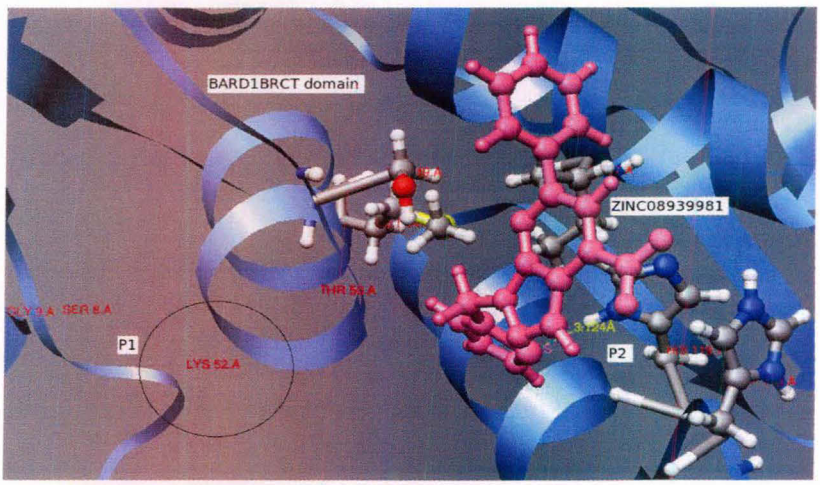


Fig 36a: ZINC Molecule (ZINC08939981) when docked in BARD1- BRCT P2 pocket is ranked 1 by Glide XP. [Major interacting residues of both protein and ligand are shown in ball & stick]

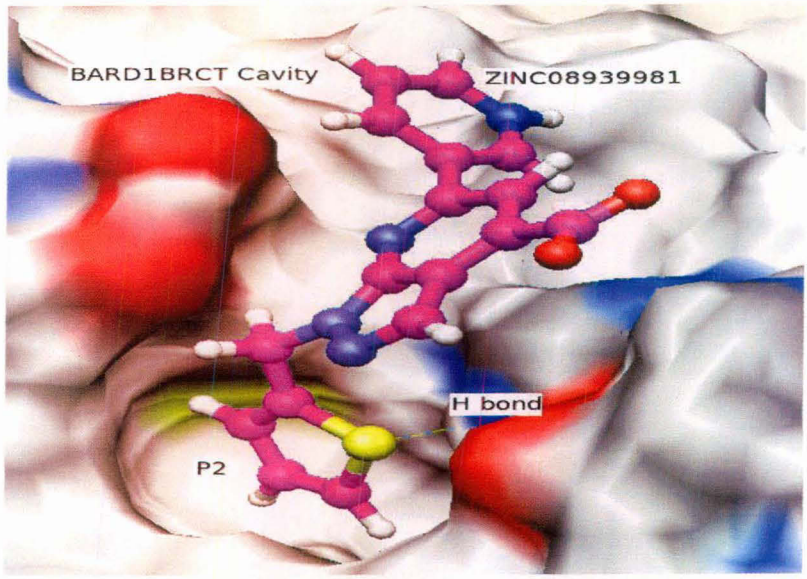


Fig 36b) Molecule with ZINC ID 08939981 ranked 1 for P2 pocket of BARD1BRCT domain (top: ribbon; bottom: surface view) by Glide XP.

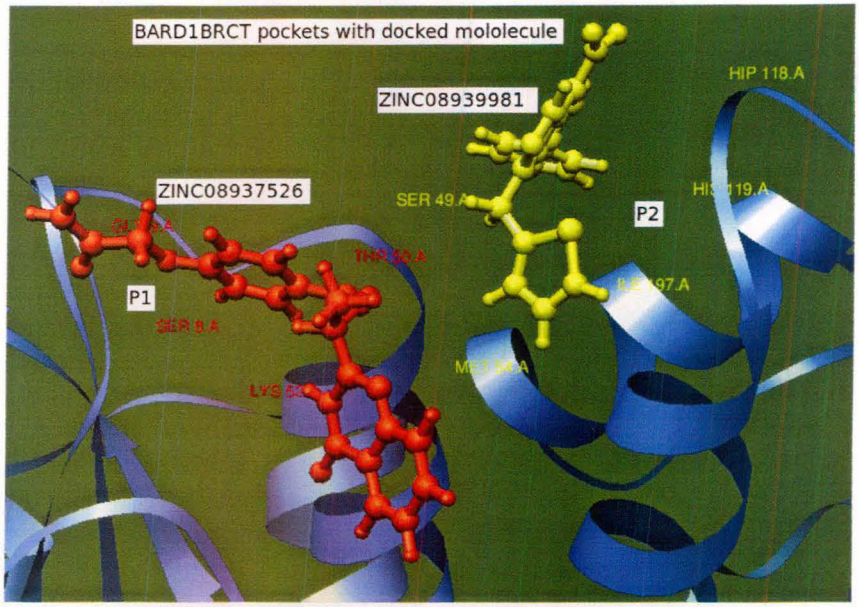
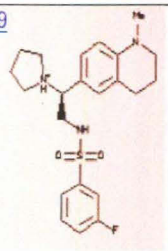


Fig 36c: P1 and P2 pocket of BARD1BRCT domain with best scored molecule ZINC ID-08937526(red) and 08939981(yellow) respectively.

Table18: Comparatively better scored molecule (ZINC ID 08799099) out of 8312 molecule of ZINC database.

ZINC ID 08799099	Tool	Rank	Characteristics
8799099 	GLIDE XP	19	Mwt 418.558 xLogP 3.75 N_h_donors 2 n_h_acceptors 5 psa 53
	FRED	26	
	DOCK	250	
	GOLD	4714	

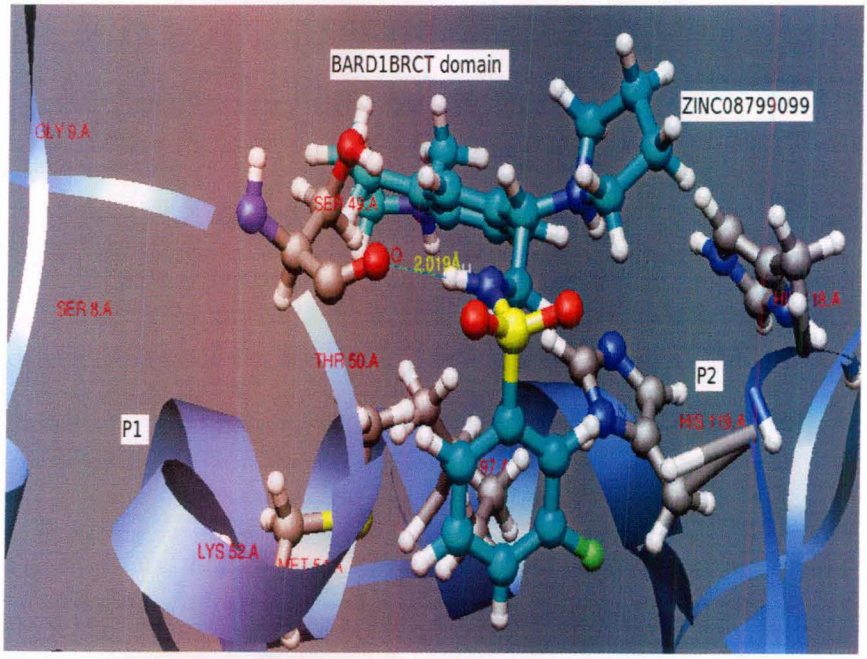


Fig 37: P2 pocket of BARD1- BRCT domain (ribbon) with molecule (ZINC ID 08799099) shows relatively better score by all four docking tools i.e. Glide XP [Rno.19], Fred [Rno.26], Dock [Rno.250] except Gold [4714]

Table 19: Interaction of top ranked ligands with their binding site [tool: Getneares]

ZINC ID	**Residues (P1 pocket)**	**Atoms**	**Distance (< 5Å)**
ZINC08937526	ILE A 573	HD11<> 6 O1	4.56
	GLY A 574	O <> 26 O5	3.75
	SER A 575	OG <> 21 C16	3.74
	GLY A 576	N <> 23 O4	3.18
	LEU A 577	O <> 27 N3	2.80
	GLN A 582	HE21 <>26O5	2.18
	GLU A 599	OE1 <> 7 C5	3.12
	PHE A 600	N <>6 O1	3.18
	THR A 617	HG21<>21C16	3.64
	LEU A 618	HD23 <> 7 C5	2.96
	LYS A 619	CG <> 6 O1	3.40
	LEU A 622	HD12 <> 6 O1	3.81
	LYS A 693	CE <>9 C7	3.81
ZINC ID	**Residues (P2 pocket)**	**Atoms**	**Distance (< 5Å)**
ZINC08939981	SER A 616	O <> 15 C12	2.85
	THR A 617	C <> 17 C14	4.04
	LEU A 618	HD13<> 18 C15	3.22
	MET A 621	CE <> 17 C14	3.23
	HIP A 685	CD2 <>22 O1	3.72
	HIS A 686	ND1 <> 20 S1	3.12
	PRO A 687	CD <> 19 C16	4.19
	ASN A 690	OD1 <> 19 C1	3.03
	LEU A 691	HD23 <>19 C16	4.95
	LEU A 694	HD11 <>18 C15	3.26
	SER A 761	O <> 23 N1	4.73
	ILE A 764	HG22 <> 4 C4	3.00
	ASP A 765	OD1 <> 23 N1	2.51
ZINC08799099*	SER A 616	O <> 13 N2	3.01
	THR A 617	CA <> 16 O2	3.74
	LEU A 618	HD13 <>19 C15	3.19
	MET A 621	CE <>18 C14	3.52
	HIP A 685	O<> 27 C22	3.72

	HIS A 686	CA <>23 F1	3.20
	PRO A 687	CD <>23 F1	3.61
	ASN A 690	CB <>23 F1	3.36
	LEU A 691	HD21 <>23 F1	3.82
	LEU A 694	HD11 <>20C16	2.87
	SER A 761	O <>28 N1	4.92
	ILE A 764	HD11 <>20 C16	3.08
	ASP A 765	OD1 <> 2 C2	2.79
	MET A 768	CE <>4 C4	4.99

*Molecule with ZINC ID 08799099 shows relatively better score by all four docking tools except GOLD

Table20: Ranking of top scored molecule by Fred & comparison with other tools

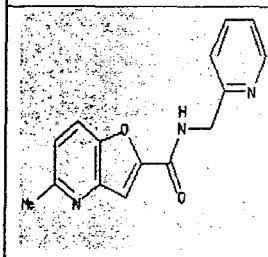
ZINC ID 08798570	Tool	Rank	Characteristics
	FRED	1	Mwt 267.288
	GLIDE	229	xLogP 1.08
	DOCK	2000+	N_h_donors 1 n_h_acceptors 5
	GOLD	5205	psa 68

Table 21: Ranking of top scored molecule by DOCK & comparison with other tools

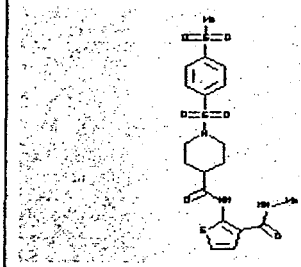
ZINC ID 08933378	Tool	Rank	Characteristic
	DOCK	1	Mwt 485.069
	GLIDE XP	755	xLogP 0.6
	FRED	1000+	N_h_donors 2 n_h_acceptors 9
	GOLD	3218	psa 129

Table 22: Ranking of top scored molecule by DOCK & comparison with other tools

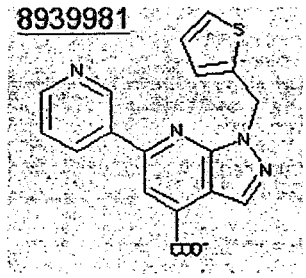
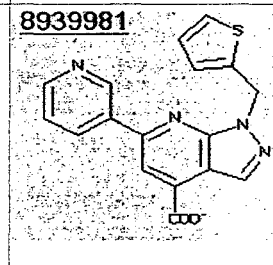
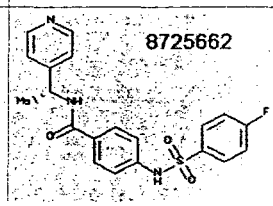
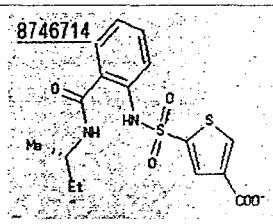
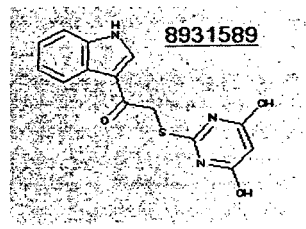
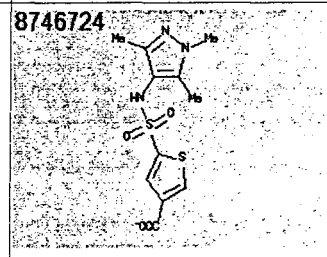
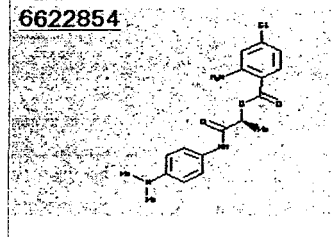
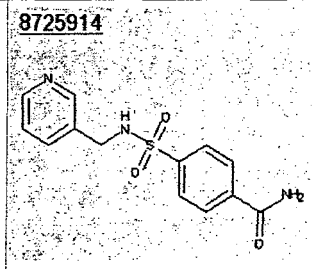
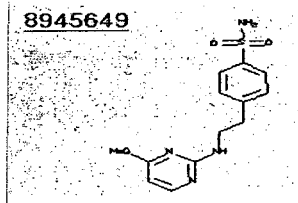
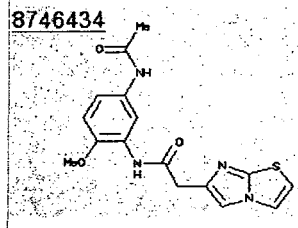
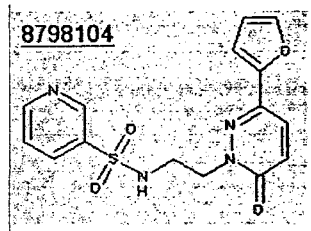
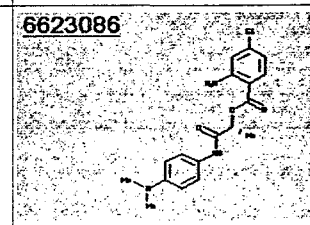
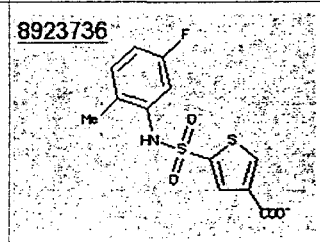
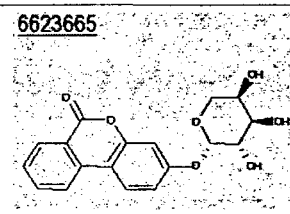
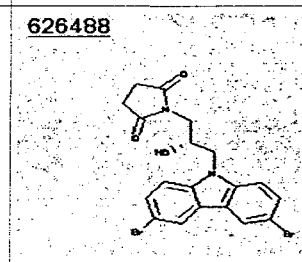
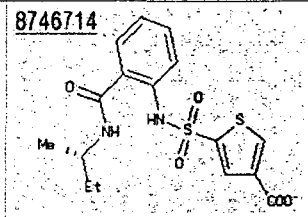
ZINC ID 08939981	Tool	Rank	Characteristic
	GLIDE XP	1	MWT 335.368 xLogP 2.61 N_h_donors 0 n_h_acceptors 6 psa 83
	DOCK	2000+	
	FRED	1000+	
	GOLD	5641	

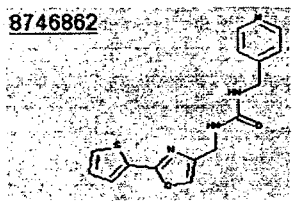
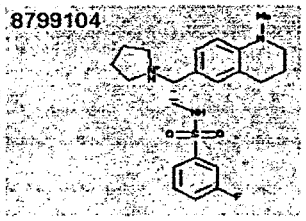
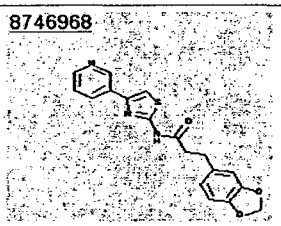
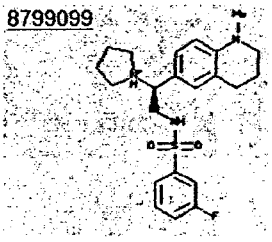
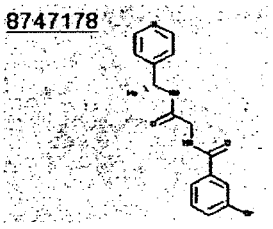
Table 23: Rank, ZINC ID/Structure, Score and Characteristics of top 20 Gscore.

[Note: Rank above 2000 for GridScore (DOCK) & above 1000 for ChemGauss3 (Fred) are not mentioned, psa; polar surface area]

Rank	ZINC ID	Score	Characteristic
GLIDE XP 1 GOLD 5641 DOCK FRED		G Score -6.24 Gold Score 39.56	MWT 335.368 xLogP 2.61 N_h_donors 0 n_h_acceptors 6 psa 83
GLIDE XP 2 GOLD 4250 DOCK FRED		G Score -6.18 Gold Score 42.76	MWT 399.447 xLogP 2.73 N_h_donors 2 n_h_acceptors 6 psa 88
GLIDE XP 3 GOLD 997 DOCK FRED		G Score -6.08 Gold Score 50.2	MWT 381.455 xLogP 2.51 N_h_donors 2 n_h_acceptors 7 psa 115

GLIDE XP 4 GOLD 4788 DOCK FRED		G Score -5.64 Gold Score 41.63	MWT 301.327 xLogP 2.81 N_h_donors 3 n_h_acceptors 6 psa 99
GLIDE XP 5 GOLD 4505 DOCK FRED		G Score -5.57 Gold Score 42.27	MWT 314.368 xLogP 0.97 N_h_donors 1 n_h_acceptors 7 psa 104
GLIDE XP 6 GOLD 168 DOCK FRED		G Score -5.5 Gold Score 54.66	MWT 361.829 xLogP 3.65 N_h_donors 3 n_h_acceptors 6 psa
GLIDE XP 7 GOLD 4318 DOCK FRED		G Score -5.48 Gold Score 42.65	MWT 291.332 xLogP -0.01 N_h_donors 3 n_h_acceptors 6 psa 102
GLIDE XP 8 GOLD 4710 DOCK FRED		G Score -5.37 Gold Score 41.82	MWT 308.363 xLogP 1.29 N_h_donors 3 n_h_acceptors 7 psa 107
GLIDE XP 9 GOLD 3856 DOCK1359 FRED		G Score -5.33 Gold Score 43.65 Grid Score -32.59	MWT 344.396 xLogP 1.46 N_h_donors 2 n_h_acceptors 7 psa 84

GLIDE XP 10 GOLD 2530 DOCK FRED	8798104 	Gscore -5.33 Gold Score 46.38	MWT 418.558 xLogP 3.75 N_h_donors 2 n_h_acceptors 5 psa 53
GLIDE XP 11 GOLD 4778 DOCK FRED	6623086 	Gscore -5.27 Gold Score 41.64	MWT 361.829 xLogP 3.65 N_h_donors 3 n_h_acceptors 6 psa 84
GLIDE XP 12 GOLD 5816 DOCK FRED	8923736 	Gscore -5.25 Gold Score 39.06	MWT 314.339 xLogP 2.69 N_h_donors 1 n_h_acceptors 5 psa 86
GLIDE XP 13 GOLD 6332 DOCK 346 FRED	6623665 	Gscore -5.23 Goldscore 27.29 Grid Score -11.72	MWT 344.319 xLogP 1.26 N_h_donors 3 n_h_acceptors 7 psa 109
GLIDE XP 14 GOLD 5351 DOCK FRED	626488 	Gscore -5.22 Goldscore 40.37	MWT 480.156 xLogP 3.54 N_h_donors 1 n_h_acceptors 5 psa 62
GLIDE XP 15 GOLD 997 DOCK FRED	8746714 	Gscore -5.17 Goldscore 50.2	MWT 381.455 xLogP 2.51 N_h_donors 2 n_h_acceptors 7 psa 115

GLIDE XP 16 GOLD 4208 DOCK FRED	8746862 	Gscore -5.17 Goldscore 42.86	MWT 314.37 xLogP 1.49 N_h_donors 2 n_h_acceptors 6 psa 80
GLIDE XP 17 GOLD 1480 DOCK FRED	8799104 	Gscore -5.06 Goldscore 48.74	MWT 418.558 xLogP 3.75 N_h_donors 2 n_h_acceptors 5 psa 53
GLIDE XP 18 GOLD 4388 DOCK FRED	8746968 	Gscore -5.03 Goldscore 42.49	MWT 353.403 xLogP 2.98 N_h_donors 1 n_h_acceptors 6 psa 73
GLIDE XP 19 **GOLD 4714** **DOCK 250** **FRED 26**	8799099 	Gscore -5.03 Goldscore 41.8 Grid Score -35.07 Fred (Chemgauss3) 22.28	MWT 418.558 xLogP 3.75 N_h_donors 2 n_h_acceptors 5 psa 53
GLIDE XP 20 GOLD 3911 DOCK FRED	8747178 	Gscore -5.03 Goldscore 43.52	MWT 362.227 xLogP 2.19 N_h_donors 2 n_h_acceptors 5 psa 71

CHAPTER 8.

DISCUSSIONS & CONCLUSIONS

Many comparative studies have been reported in the past decades that evaluate the relative performance of the most popular docking programs [Bissantz C et al., 2000, Stahl M and Rarey M, 2001, Perez C and Ortiz A.R,2001, Kellenberger E et al., 2004, Parola E et al., 2004,]. In this study the docking tools DOCK, GOLD, GLIDE and FRED which is widely used in industries as well as academia have been compared for docking of 8132 drug-like molecules from ZINC database. The target protein BARD1- BRCT domain (2NTE.pdb) [Birrane G,2007] has no known bound ligand, hence optimization of parameters for docking protocol was done taking reference of T4 Lysozyme bound with known inhibitor Isobutyl Benzene (pdb id 184L.pdb).

Using active site detection program CASTp (Computed Atlas of Surface Topography of proteins) putative binding pockets of BARD1- BRCT domain have been identified and verified. These pockets correspond to pocket P2 and P1 as described in literature [Birrane G et al., 2007]. P2 pocket having two prominent residues His 685 & His 686 is ranked one by CASTp. Docking has been performed for putative binding pocket P2 by all four docking tools while molecules were docked in P1 pocket using docking tool Glide.

All the 8132 molecule of a subset of molecule from ZINC database were ranked by Glide XP of Glide and Goldscoe of GOLD, while only top 2000 & 1000 molecules were ranked by GridScore of DOCK and Chemgauss3 of FRED respectively. To get the true positive cases top 50 best scored (Glide XP) molecules were ranked and compared to ranking based on scoring function by other docking tools. The comparision shows wide variation in ranking even for the same molecules (table 16). Molecule with ZINC ID 08799099 was found to score relatively better by all the software except GOLD. This molecule has been ranked- 19 by GLIDE XP, 26 by FRED, 250 by DOCK and 4714 by GOLD. Poor performance by GOLD may be due to small Genetic Algorithm run per ligand which is kept 10 in this study was too low

a number to be used. The overall variation in results may be due to different algorithm used for searching and scoring by different docking tools, different way to define the binding site etc. For example in GOLD; the input is the atom number of one of the active residue while in Glide; list of active residue were provided. For DOCK, the active site was defined taking reference of the ligand from bound ligand-protein complex which showed top G Score using Glide SP. Since the target protein BARD1-BRCT domain doesn't contain any known bound ligand, molecule showing topmost GScore by Glide SP (i.e. ZINC08928542) was taken as reference. Clusters were examined for the target, and the cluster covering the binding site (or active residues) was chosen by selecting those spheres within 10 Å of the bound protein-ligand complex.

Conclusions

On the basis of above study it is concluded that consideration of false negative cases taking lowest ranking molecules from the study, using larger dataset from library, more than one target protein as case study and inclusion of more parameters other including RMSD and using a consensus scoring function would provide better insight to the comparative study of different tools. Moreover, highly conserved BRCT domain of other proteins which are well known to be bound with phospho-peptide (e.g. BRCA1-Ct1p, BRCA1-BACHI, MDC1-γ H2AX etc) provide clue for docking of pocket P1 of BARD1-BRCT against phospho-peptide library selectively.

Limitations of the study

The electrostatic potential of the P2 pocket (Fig38) is dramatically altered by changes in the protonation state of His685 and His686. At near neutral pH, the solvent exposed $N^{\delta 1}$ and $N^{\varepsilon 2}$ atoms of His685 and His686 are not protonated, hence the P2 pocket has a negative electrostatic potential ,whereas at more acidic pH, these atoms gets protonated, switching the net charge to positive. This phenomenon is referred as "Histidine Switch".

Histidine Switch raises the possibility that the BARD1-BRCT interaction with its ligand/s may be dramatically regulated by the protonation of His685 and His686 in response to pH shift in local cellular environments during various physiological and

pathological conditions [Boyer, M. J et al., 1996 & Lee, D et al., 2006].

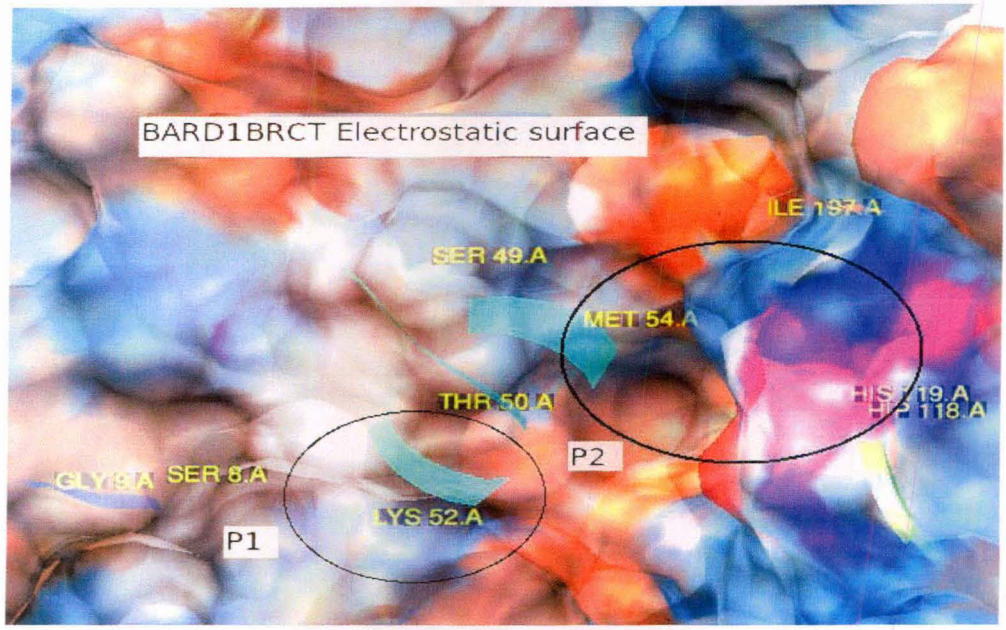


Fig38: Electrostatic potential Surface of BARD-BRCT domain (tool:Chimera).

From perspective of docking studies "Histidine Switch" marks major limitation as protonation state of macromolecule is not considered.

REFERENCES

1. Abagyan R, Totrov M, Kuznetsov D ICM—a new method for protein modeling and design: applications to docking and structure prediction from the distorted native conformation J Comput Chem 1994b, 15: 488–506
2. Aravind, L., Walker, D. R., and Koonin, E. V. Conserved domains in DNA repair proteins and evolution of repair system, Nucleic Acids Res. 1999, 27, 1223–1242
3. Bajorath J, Integration of virtual and high throughput screening, Nature Publishing Group, 2002, vol 1, 882-894.
4. Baxtar CA et al., Flexible docking using Tabu search and empirical estimate of binding affinity, Protein, 1998, 33,367-382
5. Binkowski. T.A et al., CASTp: Computed Atlas of Surface Topography of proteins, Nucleic Acids Research, 2003, Vol. 31, No. 13, 3352-3355
6. Birrane G, Verma A K, Soni A, Ladias J, Crystal structure of the BARD BRCT Domains, Biochemistry, 2007, 46, 7706-7712.
7. Bissantz C et al., Protein-based virtual screening of chemical databases: 1. Evaluation of different docking/scoring combinations. J Med Chem. 2000, 43: 4759-4767.
8. Bissantz C et al., Protein-Based Virtual Screening of Chemical Databases.1.Evaluation of Different Docking/Scoring Combinations, J. Med. Chem. 2000, 43, 4759-4767.
9. Boulton, S. J. *et al. BRCA1/BARD1* orthologs required for DNA repair in *Caenorhabditis elegans*. Curr. Biol. 2004, 14, 33–39
10. Boyer, M. J and Tannock, I. F. Regulation of intracellular pH in tumor cell lines: influence of microenvironmental conditions, Cancer Res. 1992, 52, 4441-4447.
11. Brady G. P, Stouten P. F. W. Fast prediction and visualization of protein binding pockets with PASS. J. Comput.-Aided Mol. Des. 2000, 14, 383–401.
12. Brenk, R. et al. Virtual screening for submicromolar leads of tRNAguanine transglycosylase based on a new unexpected binding mode detected by crystal structure analysis. J. Med. Chem. 2003, 46, 1133–1143
13. Brooijmans N, Kuntz ID. Molecular recognition & Docking Algorithms. Annual Rev Biophysics Biomol Struct. 2003, 32, 335-373.
14. Brown JM, Giaccia AJ: The unique physiology of solid tumors: Opportunities (and problems) for cancer therapy. Cancer Res 1998; 58:1408– 1416.
15. Brown JM: The hypoxic cell: A target for selective cancer therapy. 18th Bruce F. Cain Memorial Award lecture. Cancer Res 1999; 59:5863–5870.
16. Brzovic, P. S et al., BRCA1 RING domain cancer-predisposing mutations. Structural consequences and effects on protein–protein interactions. J. Biol. Chem. 2001, 276, 41399–41406.

17. Burchenal et al. Treatment of Acute Leukemia *Pediatrics.*1956; 18: 643-660.

18. Campbell G. Mutations in human breast cancer-an overview. Journal of the National Cancer Institute, 1989; 81:1780-1786.

19. Chai Y L et al., The second BRCT domain of BRCA1 proteins interacts with p53 and stimulates transcription from the p21WAF1/CIP1 promoter. Oncogene, 1999, 18, 263-68.

20. Chang, D. et al., Protemot: prediction of protein binding sites with automatically extracted geometrical templates. Nucleic Acids Res. 2006, 34, W303–W309.

21. Chen Y et al., Computational Method for Drug Target Search and Application in Drug Discovery, Journal of Theoretical and Computational Chemistry, 2002, vol1, Part1, 213-224. www.dspace.mit.edu/bitstream/handle/1721.1/3777/MEBCS004.pdf?sequence =2

22. Clapperton et al., Structure and mechanism of BRCA1 BRCT domain recognition of phosphorylated BACH1 with implications for cancer, Nature Structural & Molecular Biology, June 2004, Volume 11, Number 6, 512-518.

23. Claußen H, et al., FLEXE: efficient molecular docking considering protein structure variations. J mol Biol., 2001, 308: 377–395.

24. Collins I & Workman P. New approaches to molecular cancer therapeutics, Nature Chemical Biology December 2006, volume 2, number 12,689-700

25. Diller D.J, High throughput docking for library design and library prioritization. Proteins 2001; 43:113–124.

26. DiMasi J A et al., The price of innovation: new estimates of drug development costs. Journal of Health Economics, 2003, 22,151–185

27. Dixon J.S, Evaluation of the CASP2 Docking Section, PROTEINS: Structure, Function, and Genetics, 1997, Suppl. 1:198–204

28. Doman, T.N. et al. Molecular docking and high-throughput screening for novel inhibitors of protein tyrosine phosphatase-1B. J. Med. Chem.2002, 45, 2213–2221

29. Duhovny et al., Predicting Molecular Interaction *in silico*: II Protein-Protein and Protein Drug Docking, Current Medicinal Chemistry, 2005, II, 91-107.

30. Dunbrack R.L et al., Meeting review: the Second Meeting on the Critical Assessment of Techniques for Protein Structure Prediction (CASP2), Asilomar, California, December 13-16, 1996; Folding and Design, 1997, 1, R27-R42

31. Dundas J et al., CASTp: computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues, Nucleic Acids Research, 2006, Vol. 34, W116–W118, Web Server issue.

32. Easton D F. How many more breast cancer predisposition genes are there? Breast cancer Res 1999; 1:14-17.

33. Edwards R.A et al., The BARD1 C-Terminal Domain Structure and Interactions with Polyadenylation Factor CstF-50 Biochemistry 2008, 47, 11446–11456

34. Eldridge M. D et al., Empirical scoring functions 1. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. J. Comput.-Aided Mol. Des. 1997, 11, 5, 425-445.

35. Elford H.L: Effect of hydroxyurea on ribonucleotide reductase. Biochem Biophys Res Commun 1968:33:129–135.

36. Esteller M, Cancer epigenomics: DNA methylomes and histone-modification maps, Nature, April, 2007, 8, 286-294

37. Ewing T.J.A, kuntz I.D, Critical Evaluation of Search Algorithms for Automated Molecular Docking and Database Screening, J Comput Chem 1997,18: 1175-1189;

38. Ewing T.J.A et al., DOCK 4.0: Search strategies for Automated Molecular Docking of flexible molecule database, Journal of Computer-Aided Molecular Design, 2001, 15: 411–428.

39. Farmer H et al., Targeting DNA repair defect in BRCA mutant cells as a therapeutics strategy. Nature, 14 April 2005, vol 434, 917-920.

40. Feki, A. *et al.* BARD1 expression during spermatogenesis is associated with apoptosis and hormonally regulated. Biol. Reprod. 2004, 71, 1614–1624

41. Ferneaux P et al., Low expression of bcl-2 in BRCA1 associated breast cancer. 2000; 83:1318-1322.

42. Finn, R. D. et al., iPfam: Visualization of protein-protein interactions in PDB at domain and amino acid resolutions. *Bioinformatics* 2005, *21*, 410–412.

43. Friesner R. A et al., Extra precision glide: Docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. J. Med. Chem. 2006, 49 (21), 6177-6196.

44. Friesner R.A et al. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. J Med Chem, 2004, 47, 1739-1749.

45. Gabdoulline, R. R et al., MolSurfer: a macromolecular interface navigator. Nucleic Acids Res. 2003, 31, 3349–3351.

46. Gautier F et al Identification of an apoptotic cleavage product of BARD1 as an autoantigen: a potential factor in the antotumoral response mediated by apoptotic bodies. Cancer res, 2000, 60, 6895-6900.

47. Geist Al et al., "PVM -Parallel Virtual Machine: A User's Guide and Tutorial for Networked Parallel Computing", MIT Press, 1994.

48. Glover J.N.M et al.,Interactions between BRCT repeats and Phosphoproteins: tangled up in two. Trends in Biochemical Sciences, 2004, Vol.29 No.11

49. Gold, N. D.; Jackson, R. M. Sitesbase: a database for structure-based protein-ligand binding site comparisons. Nucleic Acids Res. 2006, 34, D231–D234.

50. Gong, S. et al., PSIbase: a database of protein structural interactome map (PSIMAP). Bioinformatics 2005, 21, 2541–2543.

51. Goodsell D.S and Olsonn A.J, Automated Docking of Substrates to Proteins by Simulated Annealing, PROTEINS Structure, Function, and Genetics, 1990, 8:195-202

52. Grzybowski, B.A. et al. Combinatorial computational method gives new picomolar ligands for a known enzyme. Proc. Natl. Acad. Sci. USA 2002, 99, 1270–1273

53. Guda, C et al., DMAPS: a database of multiple alignments for protein structures. *Nucleic Acids Res.* 2006, *34*, D273– D276.

54. Ha Sookhee et al., Evaluation of docking/scoring approaches: A comparative study based on MMP3 inhibitors, Journal of Computer-Aided Molecular Design, 2000, 14: 435–448

55. Halgren. T.A et al., Glide: A New Approach for Rapid, Accurate Docking and Scoring. 2. Enrichment Factors in Database Screening, J. Med. Chem. 2004, 47, 1750-1759.

56. Halperin I et al., Principal of Docking: An overview of search algorithms and a guide to scoring functions, 2002, Proteins: structure, function, and genetics, 47,409-443.

57. Hanahan D, Weinberg RA. The Hallmarks of Cancer. Cell 2000; 100:50-70.

58. Hanover JA: Glycan-dependent signaling: O-linked N-acetylglucosamine. FASEB J 2001; 15:1865–1876.

59. Hartwell, L. H. & Weinert, T. A. Checkpoints: controls that ensure the order of cell cycle events. Science, 1989, 246, 629–634.

60. Hashizume R, Fukuda M, Maeda I, Nishikawa H, Oyake D, Yabuki Y, et al. The RING heterodimer BRCA1–BARD1 is a ubiquitin ligase inactivated by a breast cancer-derived mutation. J Biol Chem 2001; 276:14537–40.

61. Henkels KM, Turchi JJ, Cisplatin-induced apoptosis proceeds by caspase-3-dependent and independent pathways in cisplatin-resistant and -sensitive human ovarian cancer cell lines. Cancer Res 1999; 59: 3077–3083.

62. Hu X et al., A practical approach to docking of zinc metalloproteinase inhibitors. J Mol Graph Model, 2004, 22, 293-307.

63. Huang CH, Treat J, New advances in lung cancer chemotherapy: Topotecan and the role of topoisomerase I inhibitors. Oncology 2001; 61(suppl 1):14–24.

64. Irminger-Finger I et al., BRCA1-dependent and independent functions of BARD1. The International Journal of Biochemistry & Cell Biology 2002, 34, 582–587

65. Irwin J J, and Brian K. Shoichet, ZINC – A Free Database of Commercially Available Compounds for Virtual Screening, *J. Chem. Inf. Model.*, 2005, 45 (1), 177-182

66. Ivanisenko, V. A et al., PDBSite: a database of the 3D structure of protein functional sites. *Nucleic Acids Res.* 2005, *33*, D183–D187

67. Jain A.N, Surflex-Dock 2.1: Robust performance from ligand energetic modeling, ring flexibility, and knowledge-based search, J Comput Aided Mol Des, 2007, 21:281–306

68. Janin J et al., CAPRI: A critical assess of predicted interactions, PROTEINS: Structure, Function, and Genetics, 2003, Suppl. 52:2-9

69. Johnston SRD, Kelland LR: Farnesyl transferase inhibitors – a novel therapy for breast cancer. Endocr Relat Cancer 2001; 8:227–235.

70. Jones G et al., Development and Validation of a Genetic Algorithm for Flexible Docking, J. Mol. Biol., 1997, 267, 727-748

71. Joseph-McCarthy D et al., Pharmacophore-based molecular docking to account for ligand flexibility. Proteins Struct Funct Genet, 2003, 51: 172–188.

72. Joukov V et al., Functional communication between endogenous BRCA1 and its partner, BARD1, during *Xenopus laevis* development. Proc. Natl Acad. Sci. 2001, 98, 12078–12083.

73. Kastan M B & Bartek J, Cell-cycle checkpoints and cancer, Nature, November 2004, 432 18, 316-323

74. Kauraniamie P et al., MYC oncogene amplification in hereditary BRCA1 breast cancer. Cancer Rews 2000; 27: 247-254.

75. Kellenberger E et al., Comparative evaluation of Eight Docking Tools for Docking and Virtual Screening Accuracy PROTEINS: Structure, Function, and Bioinformatics, 2004, 57: 225–242

76. Kim HJ et al. NF- κB and IKK as therapeutic targets in cancer. Cell Death and Differentiation, 2006, 13, 738–747

77. Kirchmair,J et al., The Protein Data Bank (PDB), Its Related Services and Software Tools as Key Components for *In Silico* Guided Drug Discovery, Journal of Medicinal Chemistry, 2008, Vol. 51, No. 22,7021-7044.

78. Klebes G, Virtual ligand screening: strategies, perspectives and limitations, Drug Discovery Today. July 2006, vol 11,580-594

79. Kleiman F. E and Manley J.L. The BARD1-CstF-50 Interaction Links mRNA 39 End Formation to DNA Damage and Tumor Suppression Cell, 2001Vol. 104, 743–753.

80. Knudson AG Jr. Mutation and cancer; statistical study of Rb. Proc Natl Acad Sci USA 1971; 68: 820-3.

81. Knudson AG Jr. Retinoblastoma: a prototypic hereditary neoplasm. Semin Oncol 1978; 5:57-60.

82. Kubinyi H, Success Stories of Computer-Aided Design, in: Computer Applications in Pharmaceutical Research and Development (Ekins, S. Ed.) [Wiley Series in Drug Discovery and Development (Wang, B. Ed.)], Wiley-Interscience, 2006, 377–424.

83. Kundrotas P.J et al., PROTCOM: searchable database of protein complexes enhanced with domain-domain structures. Nucleic Acids Res. 2007, 35, D575–D579.

84. Kuntz I.D et al., A Geometric Approach to Macromolecule-Ligand Interactions J. Mol. Biol. 1982, 161,269-288

85. Lafarge, S. & Montane, M. H. Characterization of *Arabidopsis thaliana* ortholog of the human breast cancer susceptibility gene 1: *AtBRCA1*, strongly induced by gamma rays. *Nucleic Acids Res.* 2003 **31**, 1148–1155

86. Laskowski, R. A. SURFNET: a program for visualizing molecular surfaces, cavities, and intermolecular interactions. *J. Mol.* Graphics 1995, 13, 323–330.

87. Laurie, A. T. R.; Jackson, R. M. Q-SiteFinder: an energy-based method for the prediction of protein-ligand binding sites. Bioinformatics 2005, 21, 1908–1916.

88. Lavelle F, Riou JF, Laoui A, Mailliet P: Telomerase: A therapeutic target for the third millennium? Crit Rev Oncol Hematol 2000; 34:111–126.

89. Lee KS et al., Sequential activation and production of matrix metalloproteinase-2 during breast cancer progression. Clin Exp Metastasis 1996; 14:512–519.

90. Lee, D., RAP uses a histidine switch to regulate its interaction with LRP in the ER and Golgi, *Mol. Cell,* 2006 *22*, 423-430.

91. Liang J et al., Anatomy of protein pockets and cavities: Measurement of binding site geometry and implications for ligand design, Protein Science, 1998, 7, 1884-1897.

92. Lipinski C.A et al., Experimental and computational approaches to estimate solubility and permeability in drug discovery and development setting. Advanced Drug Delivery Reviews 2001, 46, 3–26

93. Llovera M et al., Involvement of prolactin in breast cancer: Redefining the molecular targets. Exp Gerontol 2000; 35:41–51.

94. Longati P et al., Receptor tyrosine kinases as therapeutic targets: The model of the MET oncogene. Curr Drug Targets 2001; 2:41–55.

95. Lyne, P.D. et al. Identification of compounds with nanomolar binding activity for checkpoint kinase-1 using knowledge-based virtual screening. J. Med. Chem. 1968, 47, 1962–1968

96. Madhusudan S. The emerging role of DNA repair proteins as predictive, prognostic and therapeutic targets in cancer, Cancer treatment reviews (2005) 31, 603–617

97. Manke,I.A, et al., BRCT repeats as phosphopeptide-binding modules involved in protein targeting. *Science,* 2003, 302, 636–639

98. McGann MR et al., Gaussian docking functions. Biopolymers, 2003, 68: 76–90.

99. McMartin, C. and Bohacek, R.S. QXP: powerful, rapid computer algorithms for structure-based drug design. J. Comput. Aided Mol. Des. 1997 11, 333–344

100. Meng E.C et al., Orientational Sampling and Rigid-Body Minimization in Molecular Docking, PROTEINS: Structure, Function, and Genetics1993 17:266-278.

101. Meza, J. E et al., Mapping the functional domains of BRCA1. Interaction of the ring finger domains of BRCA1 and BARD1. J. Biol. Chem. 1999, 274, 5659–5665.

102. Moitessier N., Englebienne P et al. Towards the development of universal, fast and highly accurate docking/scoring methods: A long way to go, British Journal of Pharmacology, 2008, 153, S7–S26

103. Morris GM et al., Automated docking using a Lamarkian genetic algorithm and an emperical binding free energy function, Journal of computational chemistry, 1998, vol 19, No. 14, 1639-62.

104. Nissink JWM, Murray C, Hartshorn M, Verdonk ML, Cole JC, Taylor R. A new test set for validating predictions of protein–ligand interaction. Proteins 2002; 49:457–471.

105. Parill A.L, Evolutionary and genetic methods in drug design, 1996, DDT Vol. 1, no. 12,514-521

106. Parola E et al., A Detailed Comparison of Current Docking and Scoring Methods on Systems of Pharmaceutical Relevance, PROTEINS: Structure, Function, and Bioinformatics, 2004, 56:235–249

107. Perez C and Ortiz A.R, Evaluation of docking functions for protein-ligand docking, 2001, J Med Chem, 44, 3768-3785

108. Pickett, S.D. et al. Discovery of novel low molecular weight inhibitors of IMPDH via virtual needle screening. Bioorg. Med. Chem. Lett. 2003, 13, 1691–1694

109. Powers, R.A. et al. Structure-based discovery of a novel, noncovalent inhibitor of Amp C beta-lactamase. Structure, 2002, 10, 1013–1023.

110. Rarey M et al., The particle concept: placing discrete water molecules during protein–ligand docking predictions. Proteins Struct Funct Genet., 1999. 34: 17–28.

111. Rarey M, Kramer B, Lengauer T, Klebe G, A fast flexible docking method using an incremental construction algorithm. J Mol Biol., 1996, 261: 470–489.

112. Rodriguez et al., Phosphopeptide binding specificities of BRCA1-COOH-terminal (BRCT) domain, The Journal of Biological Chemistry, 2003, Vol. 278, No. 52, 52914–52918.

113. Schulz-Gasch T, Stahl M, Binding site characteristics in structure-based virtual screening: evaluation of current docking tools. Mol Model, 2003, 9:47–57

114. Schulz-Gasch T, Stahl M, Scoring functions for protein–ligand interactions: a critical perspective. Drug Discovery Today, 2004, vol.1, No.3, 231-239

115. Seifert H. J. Markus, virtual High-throughput in silico screening, Biosilico, 2003, voll No. 4,143-149

116. Shiozaki et al., Structure of the BRCT Repeats of BRCA1 Bound to a BACH1 Phosphopeptide: Implications for Signaling, Molecular Cell, 2004, Vol. 14, 405–412

117. Shoichet B.K, Bodian D.L and Irwin D, Kuntz. Molecular Docking Using Shape Descriptors, Journal of Computational Chemistry, 1992 Vol. 13, No.3, 380-397

118. Spahn L et al., Interaction of the EWS NH2 terminus with BARD1 links the Ewing Sarcoma gene to a common tumor suppressor pathway. Cancer res. 2002, 62, 4583-4587.

119. Stahl M and Rarey M, Detailed analysis of scoring functions for virtual screening. J Med Chem, 2001, 44, 1035-1042

120. Stahura FL and Jürgen Bajorath J, Virtual Screening Methods that Complement HTS, Combinatorial Chemistry & High Throughput Screening, 2004, 7, 259-269 259

121. Stucki et al., MDC1 Directly Binds Phosphorylated Histone H2AX to Regulate Cellular Responses to DNA Double-Strand Breaks, Cell, December 29, 2005, 123, 1213–1226

122. Sunderberg S.A, High-throughput and ultra-high-throughput screening: solution- and cell- based approaches, Current opinion in Biotechnology, 2000, 11, 47-53

123. Szekeres T, Novotny L, New Targets and Drugs in Cancer Chemotherapy, Medicinal Principle and Practice, 2002, vol 11,117-125

124. Szyf M: The DNA methylation machinery as a therapeutic target. Curr Drug Targets, 2000, 1:101–118.

125. Taylor R.D, Jewsberry P.J & Essex J.W, A review of protein-small molecule docking methods, Journal of computer aided molecular design, 2002, 0:1-16.

126. Taylor RD, Jewsbury PJ, Essex JW. FDS: flexible ligand and receptor docking with a continuum solvent model and soft-core energy function. J Comput Chem. 2003, 24: 1637–1656.

127. Taylor, R. M et al., Role of a BRCT domain in the interaction of DNA ligase III-a with the DNA repair protein XRCC1, Curr. Biol. 1998, 8, 877–880

128. Thai, T. H. et al. Mutations in the BRCA1-associated RING domain (BARD1) gene in primary breast, ovarian and uterine cancers. Hum. Mol. Genet. 1998 7, 195–202.

129. Tollman P et al. A Revolution in R& D, How Genomics and Genetics are transforming the biopharmaceutical industry, The Boston Consulting Group, 2001.www.bcg.com/publications/files/eng_genomicsgenetics_rep_11_01.pdf.

130. Totrov M, Abagyan R. Flexible protein–ligand docking by global energy optimization in internal coordinates. Proteins Struct Funct Genet 1997, 29: 215-220.

131. Tsuzuki, M. et al. A truncated splice variant of human BARD1 that lacks the RING finger and ankyrin repeats. Cancer Lett. 2005. 233, 108–116

132. Turkson J, Jove R: STAT proteins: Novel molecular targets for cancer drug discovery. Oncogene 2000; 19:6613–6626.

133. Turner N et al., Hallmarks of BRCAness in sporadic cancer, Nature reviews, Cancer volume 4, October 2004, 1-6.

134. Vangrevelinghe, E. et al. Discovery of a potent and selective protein kinase CK2 inhibitor by high-throughput docking. J. Med. Chem. 2003, 46, 2656–2662

135. Varmus H. The New Era in Cancer Research, Science, 2006, Science 312, 1162-1165.

136. Venkatachalam C.M et al., LigandFit: a novel method for the shape-directed rapid docking to protein active sites, Journal of Molecular Graphics and Modelling, 2003,21,289-307

137. Venkitaraman A .R. Cancer susceptibility and the functions of BRCA1 and BRCA2. Cell, 2002, 108, 171–182.

138. Verdonk ML et al., Improved protein–ligand docking using GOLD. Proteins Struct Funct Genet, 2003, 52: 609–623.

139. Verdonk ML et al., Modeling water molecules in protein–ligand docking using GOLD. J Med Chem, 2005, 48: 6504–6515.

140. Verma et al., Structural basis for cell cycle checkpoint control by the BRCA1-CtIP complex, Biochemistry, 2005 44, 10941-10946

141. Viet M et al. Assessing energy functions for flexible docking, J Comp Chem., 1998b, 19, 1612-1622

142. Welsch PL, King MC. BRCA1 and BRCA2 and the genetics of breast and ovarian cancer. Human Mol Genet. 2001; 10:705-713.

143. Wick B, Groner B: Evaluation of cell surface antigens as potential targets for recombinant tumor toxins. Cancer Lett 1997, 118,161–172.

144. Williams R. S et al., Crystal structure of the BRCT repeat region from the breast cancer associated protein BRCA1, Nature Structural Biology, 2001,Vol. 8, No. 10, 838-842.

145. Williams R. S et al., Structural basis of phosphopeptide recognition by the BRCT domain of BRCA1, 2004, Nat. Struct. Mol. Biol. 11, 519-525

146. Wu, J. Y. et al. Aberrant expression of BARD1 in breast and ovarian cancers with poor prognosis. Int. J. Cancer, 2006, 118, 1215–1226.

147. Wu, L. C et al., Identification of a Ring protein that can interact in vivo with the BRCA1 gene product Nat. Genet. 1996, 14, 430–440.

148. Wu, L. C. et al. Identification of a RING protein that can interact in vivo with the BRCA1 gene product. Nature Genet, 1996 14, 430–440.

149. Wyss, P.C. et al. Novel dihydrofolate reductase inhibitors. Structure-based vs. diversity-based library design and high-throughput synthesis and screening. J. Med. Chem. 2003, 46, 2304–2312

150. Xu, B., Kim, S. & Kastan, M. B. Involvement of Brca1 in S-phase and G (2)-phase checkpoints after ionizing irradiation. Mol. Cell Biol. 2001, 21, 3445–345. .

151. Yamane K et al. The BRCT Regions of Tumor Suppressor BRCA1 and of XRCC1 Show DNA End Binding Activity with a Multimerizing Feature. Oncogene, 1999, 18, 5194–5203

152. Yu, X et al., The C-terminal (BRCT) Domains of BRCA1 Interact in Vivo with CtIP, a Protein Implicated in the CtBP Pathway of Transcriptional Repression, J. Biol.Chem. 1998, 273, 25388–25392.

153. Zavodszky MI, et al., Distilling the essential features of a protein surface for improving protein–ligand docking, scoring, and virtual screening. J Comput Aided Mol Des 2002; 16:883–902.

154. Zhou Z et al., 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, 1599-1608

Websites

http://www-dep.iarc.fr/globocon/globocon.html

http://www-canceratlasindia.org/PriliminaryPages_1.html.

http://www.bcg.com

http://www.dock.compbio .ucsf.edu

http://www. Biosolviet.com

http://www.eyesopen.com

http://www.schroedinger.com

http://www.ccdc.cam. ac.uk

http://www.Predictioncenter.org/casp8

http://www. sanger. uk.ac/ Genetics/ CGP/census/

http://zinc.docking.org

http://cast.engr.uic.edu/cast

http://www.rcsb.org

http://dspace.mit.edu/bitstream/handle/1721.1/3777/MEBCS004.pdf?sequence=2

http://www.bcg.com/publications/files/eng_genomicsgenetics_rep_11_01.pdf