Chapter-1 Introduction

1.1 Cancer

Cancer is a class of diseases characterized by out-of-control cell growth. There are over 100 different types of cancer, and each is classified by the type of cell that is initially affected.

Cancer harms the body when damaged cells divide uncontrollably to form lumps or masses of tissue called tumors (except in the case of leukemia where cancer prohibits normal blood function by abnormal cell division in the blood stream). Tumors can grow and interfere with the digestive, nervous, and circulatory systems and they can release hormones that alter body function. Tumors that stay in one spot and demonstrate limited growth are generally considered to be benign [1].

Cancer is ultimately the result of cells that uncontrollably grow and do not die. Normal cells in the body follow an orderly path of growth, division and death. Programmed cell death is called apoptosis, and when this process breaks down, cancer begins to form. Unlike regular cells, cancer cells do not experience programmatic death and instead continue to grow and divide. This leads to a mass of abnormal cells that grows out of control.

Cells can experience uncontrolled growth if there are damages or mutations to DNA, and therefore, damage to the genes involved in cell division. Four key types of gene are responsible for the cell division process: oncogenes tell cells when to divide, tumor suppressor genes tell cells when not to divide, suicide genes control apoptosis and tell the cell to kill itself if something goes wrong, and DNA-repair genes instruct a cell to repair damaged DNA [2].

Cancer occurs when a cell's gene mutations make the cell unable to correct DNA damage and unable to commit suicide. Similarly, cancer is a result of mutations that inhibit oncogene and tumor suppressor gene function, leading to uncontrollable cell growth.

Cancer is not a single disease – it is many diseases. Yet, all cancer cells share one characteristic: Like weeds in a garden, they reproduce rampantly, crowding out the healthy cells that contribute to the functioning of our organs – our lungs, our livers, our brains.

People unfamiliar with science's day-to-day workings will sometimes wonder why scientists do not simply go into a laboratory and emerge a couple of months later with a cure for cancer. While clinical research aims directly at finding and testing new treatments to weed out cancerous cells, basic research – such as the studies conducted at the Salk Institute – strives to understand the genetic and molecular mechanisms at the heart of this process, the genes and proteins that regulate normal cell growth but become corrupted in cancer cells [3].

For the last 40 years, Salk scientists have been asking very basic questions whose answers changed the course of cancer research forever. Time and again, the practical applications of understanding the fundamentals of how our cells work have led to major advances in diagnosing and treating many different types of cancer.

Cancer is a group of diseases characterized by unregulated division and spread of cells. The cancerous cells may occur in liquids, as in leukemia. Most, however, occur in solid tumors that originally appear in various tissues in various parts of the body. By their original locations they are classified into various types of cancer, such as lung, colon, breast, or prostate cancer. Localized tumors can be removed by surgery or irradiation with high survival rates. As cancer progresses, however, it metastasizes – invading the surrounding tissues, entering the blood stream, spreading and establishing colonies in distant parts of the body. Only a third of patients with metastasized cancer survive more than five years. Invasive distensions spreading crab-like from a tumor in the breast were described by Hippocrates [4].

1.2 Carcinogens

A complex event usually involves many causal factors, which in turn are causally linked to other factors. In talking about cause and effect, we customarily designate as its cause one or a few salient factors most directly connected to its salient effects, noting a few other factors as indirect causes, relegating many factors as background conditions, and ignoring factors too vague to determine. For example, when arson investigators decide that a fire was caused by sparks from an exposed electric wire, they treat as background conditions the presence of oxygen and flammable materials nearby. What made the wire exposed they regard as indirect

causes, which may interest criminal investigators. Saliency and relevancy are context dependent**.** In cancer research, two scientific approaches operate in two general kinds of context. Epidemiology focuses on causal factors on the levels of people and population, with results that are more useful for disease prevention. Molecular cell biology focuses on causal mechanisms on the levels of genes and cells, with results that are more useful to treatment and cure. To biologists, factors identified by epidemiology are indirect causes in the mechanism of cancer development. For most people, however, these factors are the only cancer causes that they care about; they can do something about them [5].

Back in 1775, a London physician, noticing that boys who worked as chimney sweeps were more prone than average to develop cancer later in life, surmised that the disease was provoked by soot particles. Astute clinical observations such as this provide clues to cancer-inducing conditions. However, case reports alone are seldom sufficient in establishing causation; they cannot tease apart tangling factors to pinpoint causes.

For stronger evidence of causal associations, epidemiologists design careful studies to eliminate biases and zero in on crucial factors. Such projects often take a long time and involve many subjects. A famous study that started in 1951 questioned some 40,000 British physicians about their ages and smoking histories. Then it followed them for more than two decades, recorded their changing smoking habits and health conditions, and calculated statistical correlations between the number of cigarettes smoked and lung cancer incidences. This and other epidemiological studies succeeded to defeat the tobacco industry and establish a strong causal association between smoking and lung cancer.

To evaluate the causal consequences of a particular factor, epidemiologists often compare groups whose members are as similar in everything else as possible. For instance, patterns of cancer incidences vary greatly across countries. Natives of Japan had high rates of stomach but low rates of colon cancer. Natives of Hawaii had high rates of colon but low rates of stomach cancer. Was the difference caused mainly by genetic or cultural factors? The second generation of Japanese immigrants in Hawaii exhibited the cancer pattern not of their ancestors but of native Hawaiians. Such studies of immigrants reveal that most variations in cancer patterns are not inherited [6].

Only about one percent of cancers are unmistakably inherited. They occur in childhood. Strong genetic dispositions contribute to a small portion of adult cancers, (see statement D earlier in the chapter). Hormone production during reproductive cycles and other internal factors can also

contribute. However, the vast majority of cancers are attributable to what people eat and inhale, how they behave, their working conditions, viruses and bacteria, and natural and artificial radiation and chemicals. These are usually called “environmental” risk factors for cancer.

This usage of “environment,” which includes diet, life style, and personal behavior, is unfortunate. It deviates from common meaning of the word and is easily abused, leading to much confusion. A government report, after carefully explaining the broad meaning of “environment,” goes on: “Unfortunately, the statements (that 60 to 90 percent of cancer is associated with the environment and therefore is theoretically preventable) were sometimes repeated with ‘environment’ used to mean only air, water, and soil pollution.” Confusion and abuse of scientific results persist.

Epidemiology identifies risk factors and establishes causal associations but cannot pin down the mechanisms by which risk factors induce the appearance of tumors. Discovering mechanisms is the font of laboratory science, which delves into more microscopic levels. Nevertheless, epidemiological results do expose salient peculiarities that must be explained by whatever mechanism. Unlike poisons that act quickly, carcinogens take effect very slowly. Hiroshima survivors developed cancer ten years after their brief exposure to intense radiation of the atomic bomb. Lung cancer incidences rose more than two decades after the widespread of smoking. The long latent periods accede to cancer being a disease of elders; it takes so long to develop.

Carcinogens are a class of substances that are directly responsible for damaging DNA, promoting or aiding cancer. Tobacco, asbestos, arsenic, radiation such as gamma and x-rays, the sun, and compounds in car exhaust fumes are all examples of carcinogens. When our bodies are exposed to carcinogens, free radicals are formed that try to steal electrons from other molecules in the body. Theses free radicals damage cells and affect their ability to function normally [8].

1.3 The symptoms of cancer

In cancer, one of the leading causes of death in the US, cells divide wildly out of control. Can tumor suppressor genes and proto-oncogenes provide clues on how to stop this process? According to the American Cancer Society, an estimated 294,120 men and 271,510 women will die of cancer in the United States in the year 2008. Table 1 shows the predicted distribution and percentages of these deaths (American Cancer Society, 2008) [9].

Table 1: Predicted Distribution of U.S. Cancer Deaths in 2008

Males: 294,120 Cancer-Related Deaths		Females: 271,530 Cancer-Related Deaths	
Lung and bronchus	**31%**	Lung and bronchus	**26%**
Prostate	**10%**	Breast	**15%**
Colon and rectum	**8%**	Colon and rectum	**9%**
Pancreas	**6%**	Pancreas	**6%**
Liver and bile duct	**4%**	Ovary	**6%**
Leukemia	**4%**	Non-Hodgkin's lymphoma	**3%**
Esophagus	**4%**	Leukemia	**3%**
Urinary bladder	**3%**	Uterine corpus	**3%**
Non-Hodgkin's lymphoma	**3%**	Liver and bile duct	**2%**
Kidney and renal pelvis	**3%**	Brain	**2%**
All other sites	**24%**	All other sites	**25%**

Cancer symptoms are quite varied and depend on where the cancer is located, where it has spread, and how big the tumor is. Some cancers can be felt or seen through the skin - a lump on the breast or testicle can be an indicator of cancer in those locations. Skin cancer (melanoma) is often noted by a change in a wart or mole on the skin. Some oral cancers present white patches inside the mouth or white spots on the tongue [10].

Other cancers have symptoms that are less physically apparent. Some brain tumors tend to present symptoms early in the disease as they affect important cognitive functions. Pancreas cancers are usually too small to cause symptoms until they cause pain by pushing against nearby nerves or interfere with liver function to cause a yellowing of the skin and eyes called jaundice. Symptoms also can be created as a tumor grows and pushes against organs and blood

vessels. For example, colon cancers lead to symptoms such as constipation, diarrhea, and changes in stool size. Bladder or prostate cancers cause changes in bladder function such as more frequent or infrequent urination.

As cancer cells use the body's energy and interfere with normal hormone function, it is possible to present symptoms such as fever, fatigue, excessive sweating, anemia, and unexplained weight loss. However, these symptoms are common in several other maladies as well. For example, coughing and hoarseness can point to lung or throat cancer as well as several other conditions [11].

When cancer spreads, or metastasizes, additional symptoms can present themselves in the newly affected area. Swollen or enlarged lymph nodes are common and likely to be present early. If cancer spreads to the brain, patients may experience vertigo, headaches, or seizures. Spreading to the lungs may cause coughing and shortness of breath. In addition, the liver may become enlarged and cause jaundice and bones can become painful, brittle, and break easily. Symptoms of metastasis ultimately depend on the location to which the cancer has spread [12].

1.4 Prevention of cancer

More than cures, scientists are cautiously optimistic about the possibility of improving early detection and prevention of cancer. Cancer takes several steps and a long time to develop (Figure 1). Its long latent period gives many opportunities to catch cells in their early stages of mutation and intervene to stop cancer progression. For instance, the Pap smear followed by surgical removal of detected lesions have reduced death rate of cervical cancer by almost 80 percent. To extend the success in cervical cancer to cancer in general, scientists strive to identify biological markers that can finger incipient cancerous cells and predict whether they will evolve to significant cancer. They hope to develop noninvasive and inexpensive ways to screen for these markers in cells or body fluids that a person discharges. Novel imaging techniques will locate the cancerous lesions in the body, so that they will be removed before metastasis. Or drugs targeted at specific regulatory defects will be developed that, if not curing cancer, would push its onset to beyond person’s natural life span. A fashionable vision in the genomic era is drugs customized to specific groups of people with certain genetic dispositions. It is especially pertinent to cancer, where genetic dispositions play heavy roles and a drug is

frequently effective only in a subset of patients. Instead of treating a common cancer such as breast cancer as a single disease, why not divide it into twenty diseases according to their different genetic defects, design drugs for each, and prescribe them to target patients identified by genetic tests [13].

Cancers that are closely linked to certain behaviors are the easiest to prevent. For example, choosing not to smoke tobacco or drink alcohol significantly lower the risk of several types of cancer - most notably lung, throat, mouth and liver cancer. Even if you are a current tobacco user, quitting can still greatly reduce your chances of getting cancer.

Skin cancer can be prevented by staying in the shade, protecting yourself with a hat and shirt when in the sun, and using sunscreen. Diet is also an important part of cancer prevention since what we eat has been linked to the disease. Physicians recommend diets that are low in fat and rich in fresh fruits and vegetables and whole grains [14].

Certain vaccinations have been associated with the prevention of some cancers. For example, many women receive a vaccination for the human papilloma virus because of the virus's relationship with cervical cancer. Hepatitis B vaccines prevent the hepatitis B virus, which can cause liver cancer.

Cancer prevention in some times is based on systematic screening in order to detect small irregularities or tumors as early as possible even if there are no clear symptoms present. Breast self-examination, mammograms, testicular self-examination, and Pap smears are common screening methods for various cancers.

The studies were designed to find correlations between smoking cannabis and cancers of the lung, throat, head and neck. Instead, the researchers discovered that the cancer rates of cannabis smokers were at worst no greater than those who smoked nothing at all or even better [15].

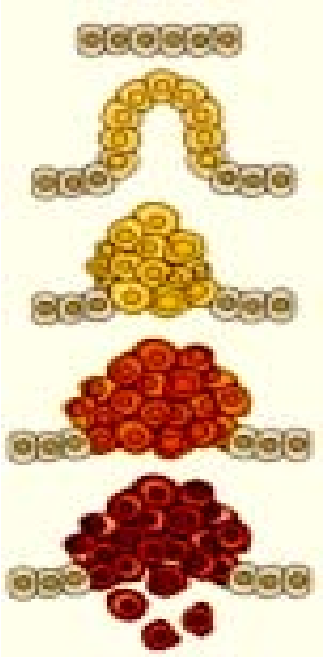


Figure 1: Multistep of Cancer Process

1.5 B-cell lymphoma 2(Bcl-2)

Bcl-2 is an intracellular protein that localizes to mitochondria, endoplasmic reticulum, and the nuclear envelop and has been shown to block apoptosis without inducing cellular proliferation. Bcl-2 is the founding member of the Bcl-2 family of apoptosis regulator proteins encoded by the BCL2 gene. Bcl-2 derives its name from B-cell lymphoma 2, as it is the second member of a range of proteins initially described in chromosomal translocations involving chromosomes 14 and 18 in follicular lymphomas (Figure 2). Bcl-2 orthologs have been identified in numerous mammals for which complete genome data are available. The two isoforms of Bcl-2, isoform 1, also known as 1G5M, and isoform 2, also known as 1G5O/1GJH, exhibit similar fold. However, results in the ability of these isoforms to bind to the BAD and BAK proteins, as well as in the structural topology and electrostatic potential of the binding groove, suggest differences in antiapoptotic activity for the two isoforms. BCL-2 is a human proto-oncogene located on chromosome 18. Its product is an integral membrane protein (called Bcl-2) located in the membranes of the endoplasmic reticulum (ER), nuclear envelope, and in the outer membranes of mitochondria. The gene was discovered as the translocated locus in a B-cell leukemia (hence the name). This translocation is also found in some B-cell lymphomas [16].

In the cancerous B cells, the portion of chromosome 18 containing the BCL-2 locus has undergone a reciprocal translocation with the portion of chromosome 14 containing the antibody heavy chain locus. This (14;18) translocation places the BCL-2 gene close to the heavy chain gene enhancer [17].

This enhancer is very active in B cells (whose job it is to synthesize large amounts of antibody). So it is not surprising to find that the Bcl-2 protein is expressed at high levels in these t(14;18) cells.

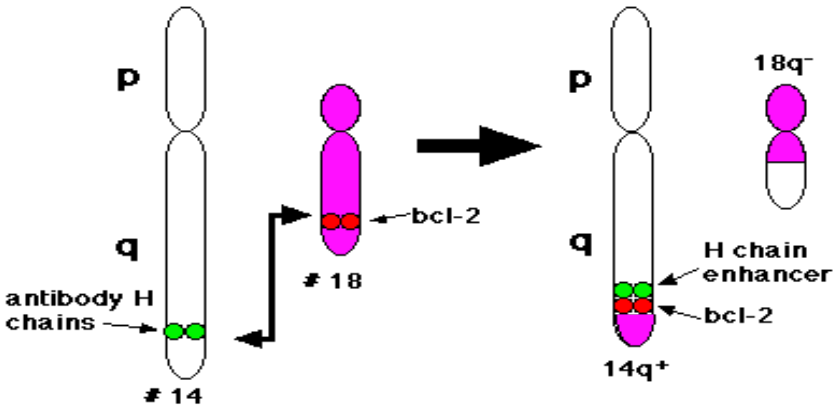


Figure 2: Locus chromosomes 14 and 18 containing the BCL-2

Some basic features of BCL-2

The Bcl-2 gene codes for a 25 kDa protein. The C terminal 21 amino acids encode a stretch of hydrophobic amino acids that are important in membrane docking: Bcl-2 resides on the cytoplasmic face of the mitochondrial outer membrane, the nuclear envelop, and the endoplasmic reticulum. Deletion of the C terminus does not abrogate Bcl-2 survival function. Most Bcl-2 homologs have this hydrophobic C terminal domain, though they not necessarily are located on membranes but are cytosolic [18].

Chapter-2 Review of Literature

When homologs of Bcl-2 have been identified, it became apparent that the Bcl-2 family can be defined by the presence of conserved motifs known as Bcl-2 homology domains (BH1 to BH4). While Bcl-2 and its most similar prosurvival homologs Bcl-XL and Bcl-w contain all four BH domains, the other pro-survival members contain at least BH1 and BH2. While the members of the Bax subfamily contain BH1, BH2 and BH3, and resemble Bcl-2 fairly closely, the seven mammalian members of the BH3 subfamily possess only the central short (9 - 16 residue) BH3 domain and are unrelated to any known protein. The BH3 subfamily members may well represent the physiological antagonists of the pro-survival proteins, since programmed cell death in *C. elegans* requires EGL-1 (the one non-mammalian BH3 family member) and Bid was reported to link caspase-8 activity to cytochromec release [19].

Pro- and anti-apoptotic family members can heterodimerize: the BH1, BH2 and BH3 domains of an anti-apoptotic member (e.g. Bcl-XL) form a hydrophobic cleft to which a BH3 amphipathic alpha-helix can bind (Sattler et al., 1997, Science, 275: 983). This BH3 cleft coupling, reminiscent of ligand-receptor engagement, may account for all dimerization within the family. Heterodimerization is not required for pro-survival function (contrary to early indications) but is essential for the pro-apoptotic activity in the BH3 subfamily. The Bax subfamily members do not depend essentially on heterodimerization, but possibly have an independent cytotoxic impact [20].

2.1 What makes BCL-2 a proto-oncogene?

B cells, like all activated lymphocytes, die a few days after they have had a chance to do their job. This ensures that they do not linger around after the threat has been dealt with and turn their attack against self components. Aging B cells kill themselves by apoptosis.

But high levels of the Bcl-2 protein protect the cells from early death by apoptosis. The Bcl-2 protein suppresses apoptosis by preventing the activation of the caspases that carry out the process [21].

So genes encoding inhibitors of apoptosis must be added to the list of genes that can act as oncogenes. In this case the effect is not achieved by increasing the rate of cell proliferation but by reducing the rate of cell death.

Although the (14:18) translocation is found in B-cell lymphomas and leukemias, something else must contribute to creating the cancer because over 50% of us have small numbers of B-cells with that translocation that never progress to cancer [22].

The antibody gene loci are dangerous places for proto-oncogenes to take up residence. Translocation of the proto-oncogene c-myc close to the enhancer of the antibody heavy chain genes also produces cancerous B cells resulting in Burkitt's lymphoma.

The translocation of the BCL-2 locus is just one of many mutations that can give rise to a malignant clone of B cells. All of the resulting leukemias are designated chronic lymphocytic leukemia or CLL [23].

2.2 Role of BCL-2 in disease

The Bcl-2 gene has been implicated in a number of cancers, including melanoma, breast, prostate, and lung carcinomas, as well as schizophrenia and autoimmunity. It is also thought to be involved in resistance to conventional cancer treatment. This suggests that decreased apoptosis may play a role in the development of cancer.

Cancer occurs when the homeostatic balance between cell growth and cell death is disturbed. Research in cancer biology has discovered that a variety of aberrations in gene expression of anti-apoptotic, pro-apoptotic and BH3-only proteins can contribute to the many forms of the disease. An example can be seen in lymphomas. The over-expression of the anti-apoptotic Bcl-2 protein in lymphocytes alone does not cause cancer. But simultaneous over-expression of Bcl-2 and the proto-oncogene may produce aggressive B-cell malignancies including lymphoma. In follicular lymphoma, a chromosomal translocation commonly occurs between the fourteenth and the eighteenth chromosomes—t(14;18) — which places the Bcl-2 gene next to the immunoglobulin heavy chain locus. This fusion gene is deregulated, leading to the transcription of excessively high levels of Bcl-2. This decreases the propensity of these cells for undergoing apoptosis [24].

Apoptosis also plays a very active role in regulating the immune system. When it is functional, it can cause immune unresponsiveness to self-antigens via both central and peripheral tolerance. In the case of defective apoptosis, it may contribute to etiological aspects of autoimmune diseases. The autoimmune disease, type 1 diabetes can be caused by defective apoptosis, which leads to aberrant T cell AICD and defective peripheral tolerance. Due to the fact that dendritic cells are the most important antigen presenting cells of the immune system, their activity must be tightly regulated by such mechanisms as apoptosis. Researchers have found that mice containing dendritic cells that are Bim -/-, thus unable to induce effective apoptosis, obtain autoimmune diseases more so than those that have normal dendritic cells. Other studies have shown that the lifespan of dendritic cells may be partly controlled by a timer dependent on anti-apoptotic Bcl-2 [25].

Apoptosis plays a very important role in regulating a variety of diseases that have enormous social impacts (Figure 3). For example, schizophrenia is a neurodegenerative disease that may result from an abnormal ratio of pro- and anti-apoptotic factors. There is some evidence that this defective apoptosis may result from abnormal expression of Bcl-2 and increased expression of caspase-3.

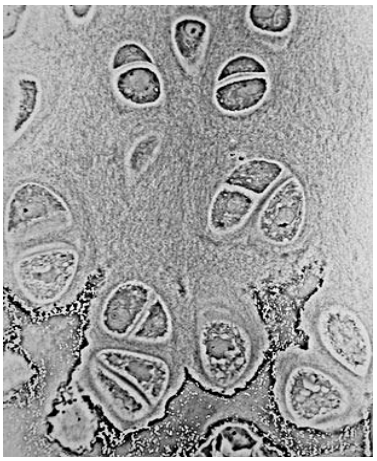


Figure 3: Apoptosis increasing from normal cells (top) to apoptotic ones (bottom)

Further research into the family of Bcl-2 proteins will provide a more complete picture on how these proteins interact with each other to promote and inhibit apoptosis. An understanding of

the mechanisms involved may help develop new therapies for treating cancer, autoimmune conditions, and neurological diseases [26].

2.3 Diagnostic use

Antibodies to Bcl-2 can be used with immuno-histochemistry to identify cells containing the antigen. In healthy tissue, these antibodies will react with B-cells in the mantle zone, as well as some T-cells. However, there is a considerable increase in positive cells in follicular lymphoma, as well as many other forms of cancer. In some cases, the presence or absence of Bcl-2 staining in biopsies may be significant for the patient's prognosis or likelihood of relapse.

The mitochondria are essential to multicellular life. Without them, a cell ceases to respire aerobically and quickly dies, a fact exploited by some apoptotic pathways. Apoptotic proteins that target mitochondria affect them in different ways (Figure 4). They may cause mitochondrial swelling through the formation of membrane pores, or they may increase the permeability of the mitochondrial membrane and cause apoptotic effectors to leak out. These are very closely related to intrinsic pathway, and tumors arise more frequently through intrinsic pathway than the extrinsic pathway because of sensitivity. There is also a growing body of evidence indicating that nitric oxide is able to induce apoptosis by helping to dissipate the membrane potential of mitochondria and therefore make it more permeable. A research done in 1999 exhibits how NO can both initiate and inhibit apoptosis due to the cellular variables [27].

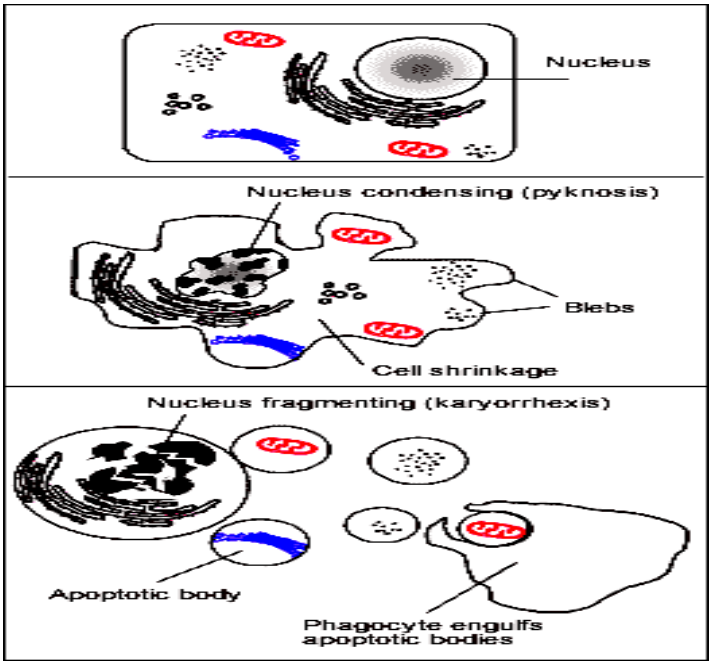


Figure 4: The role of mitochondria on process of apoptosis

Cytochromec is also released from mitochondria due to formation of a channel, the mitochondrial apoptosis-induced channel (MAC), in the outer mitochondrial membrane and serves a regulatory function as it precedes morphological change associated with apoptosis. Once cytochrome c is released it binds with Apoptotic protease activating factor - 1 (*Apaf-1*) and ATP, which then bind to *pro-caspase-9* to create a protein complex known as an apoptosome. The apoptosome cleaves the pro-caspase to its active form of caspase-9, which in turn activates the effector caspase-3 [28, 29].

A cell undergoing apoptosis shows a characteristic morphology:

1. Cell shrinkage and rounding are shown because of the breakdown of the proteinaceous cytoskeleton by caspases.
2. The cytoplasm appears dense, and the organelles appear tightly packed.
3. Chromatin undergoes condensation into compact patches against the nuclear envelope (also known as the perinuclear envelope) in a process known as pyknosis, a hallmark of apoptosis.
4. The nuclear envelope becomes discontinuous and the DNA inside it is fragmented in a process referred to as karyorrhexis. The nucleus breaks into several discrete *chromatin bodies* or *nucleosomal units* due to the degradation of DNA.

5. The cell membrane shows irregular buds known as blebs.

6. The cell breaks apart into several vesicles called *apoptotic bodies*, which are then phagocytosed [30,31].

2.4 Bioinformatics

A new field in biology has emerged after advent of large scale sequencing of DNA. The amount of data generated by genome-sequencing projects is so enormous that it cannot be stored and analyzed using conventional methods of printed material. The only way to store and analyze such vast amounts of data is by the use of computers and Information Technology.

The use of computers to manage biological information is called bioinformatics. Bioinformatics has also been defined as the science and technology related to learning, managing, and processing biological information (Figure 5). This interdisciplinary subject requires a good understanding of molecular biology, protein chemistry, mathematics, statistics, and computer languages. The aim or bioinformatics to design, create, and use software tools to generate, store, annotate, access, and analyze data and information relating to biomolecules.

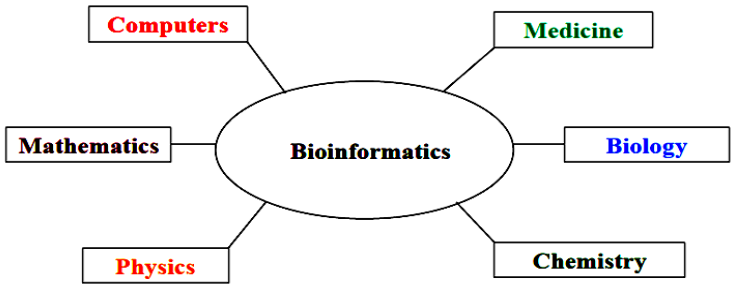


Figure 5: Bioinformatics relation with other sciences

Bioinformatics is often called computational biology, but the two fields are quite distinct. Bioinformatics uses computational and statistical techniques to consolidate, correlate, and analyze biological data. Computational biology aims to create and test mathematical models to fit experimental data in order to test a hypothesis put forward to describe a biological system [32].

2.5 The Docking

2.5.1 Drug Research at the Target Protein

Systematic research into active substances as novel drugs dates back to the second half of the 19th century. A prime example is acetylsalicylic acid, which was synthesized in 1897 by two chemists, Felix Hoffmann and Arthur Eichengrün of the company Bayer. It is now world-famous under the trade name aspirin.

It is still a disputed question as to which of the two was the actual inventor of the synthesis of acetylsalicylic acid. Regardless, this substance has neither lost its economic nor scientific importance. Since then the identification of active compounds, including those with bioactivity against infectious diseases, has been dominated by direct testing (screening) in biological systems, mostly laboratory animals. Many antibiotics in use today were discovered in the first half of the twentieth century. However, since about the 1960s the number of new drugs has steadily declined. There are a number of reasons for this, including the constant decline in the success rate of non targeted screening, the increased costs for research and development, and the higher standards of safety required. Furthermore, in the area of infectious diseases, the situation has been worsened by the emergence and increased spread of drug resistance. However, at about the same time, a new era of molecular research began in 1953 with the deciphering of the three-dimensional structure of the DNA double helix by James D. Watson and Francis H.C. Crick.

By sequencing whole genomes and the ensuing biological information, the approach to drug discovery has changed. Thus, in the target-based approach (Figure 6), in which a target protein is used to search for new active compounds, the first step is to identify those proteins that are essential to the survival of the pathogenic organism [33, 34]. The second step is to find active chemical substances that influence the isolated target protein in the desired way. Only after such optimized chemical substances with the desired activity spectrum have been found using these in vitro methods will further testing be performed in a biological system. For example, to develop a new antibiotic, an ideal prerequisite would be that the target protein is essential to the survival of the pathogenic bacteria under study and that the host organism does not also possess the same or similar protein that may also be targeted potentially resulting in toxicity. In this scenario, comparative whole genomic analysis would be well-suited to identify

Pathogen specific targets. Indeed, this approach was taken by Huynen et al. (1998) in their work on the genomes of three bacteria, *Escherichia coli, Haemophilus influenzae*, and *Helicobacter pylori*. Orthologous proteins were identified in either all three or in two of the three organisms, in addition to species-specific proteins. For *H. pylori*, the major causative agent of gastric and duodenal ulcers, the authors predicted that 123 proteins were involved in interacting between the pathogen and the host, i.e., represented potential targets for the development of an antibiotic. In pharmacological research, conserved targets usually lead to the development of broad-spectrum antibiotics, whereas with species-specific targets, narrow spectrum antibiotics are generated. Because of the increasing number of completely sequenced bacterial genomes, it is clearer which genes are generally conserved among bacteria and which are specific for certain bacterial species. However, it is not always easy to settle on the threshold of sequence similarity that blocks the pursuit of a target based drug discovery approach due to potential toxicity arising from an unwanted interaction with the human protein counterpart. For example, bacterial dihydrofolate reductase has a sequence identity of 28% at the amino acid level to the corresponding human protein, yet the antibacterial drug, trim ethoprim, is a very selective inhibitor of only the bacterial ortholog [35, 36, 37].

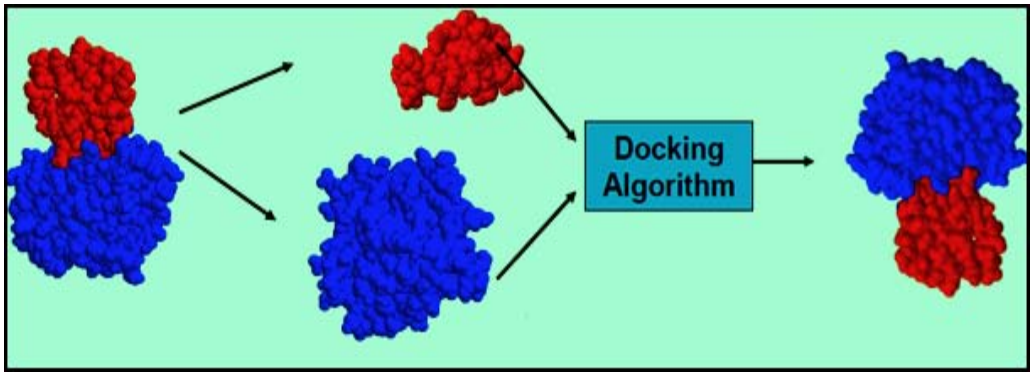


Figure 6: The docking of drug molecule and (red) to a protein receptor (blue)

The overlapping technologies of computer-assisted bioinformatics and cheminformatics have become essential components of modern drug discovery efforts. Both strategies are indispensable for the identification and validation of drug targets as well as for the screening and the design of new small-molecules. Also, of special importance is the three-dimensional structure of the drug target to allow for structure-based rational drug design. Finally, the cheminformatic approach of *virtual screening*, which tests the protein target's interaction with chemical entities in large compound libraries, is incorporated into most discovery strategies.

Unlike experiments conducted in the laboratory, virtual screening is automated being conducted at the computer, and many chemical substances can be tested for their activity spectrum relatively quickly. A variety of specialized software packages exist for structure-based rational drug design. The best known programs in use are DOCK, developed by Irvin Kuntz at the University of California, San Francisco [dock] (Ewing and Kuntz 1996) and GOLD from Peter Willett of the University of Sheffi eld [gold] as well as the programs of the Flex group developed by Thomas Lengauer and colleagues at the GMD-SCAI (now Fraunhofer-SCAI) in Sankt Augustin, Germany fl exx] (Rarey et al. 1996). The word “docking” is the modern pictorial paraphrase of the lock and key concept postulated in 1894 by Emil Fischer. The specificity of the receptor–ligand complex is brought about by the geometric complementarity of both molecules (Figures 7 and 8).

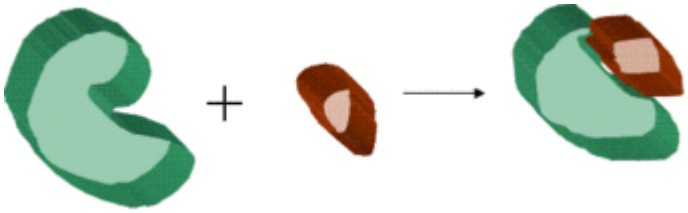


Figure 7: Schematic diagram illustrating the docking of a small molecule (brown) to a protein receptor (green) to produce a complex

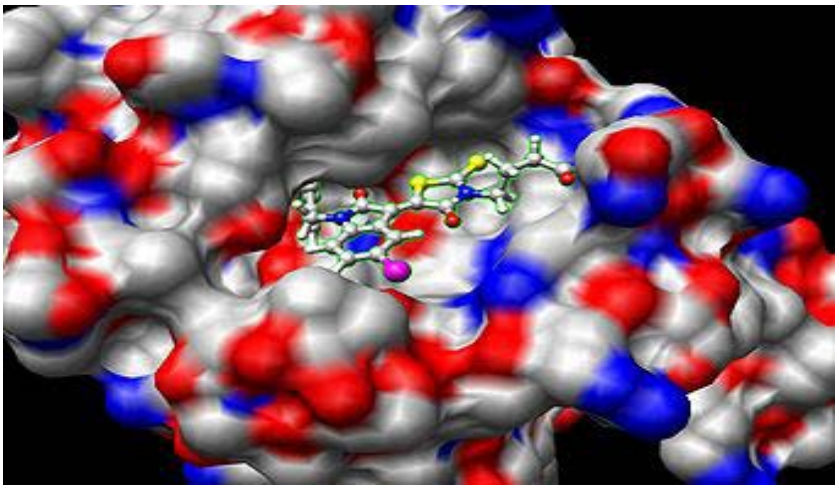


Figure 8: Small molecule docked to a protein

In the field of molecular modeling, docking is a method which predicts the preferred orientation of one molecule to a second when bound to each other to form a stable complex. Knowledge of the preferred orientation in turn may be used to predict the strength of association or binding affinity between two molecules using for example scoring functions [38].

The associations between biologically relevant molecules such as proteins, nucleic acids, carbohydrates, and lipids play a central role in signal transduction. Furthermore, the relative orientation of the two interacting partners may affect the type of signal produced (e.g., agonism vsantagonism). Therefore docking is useful for predicting both the strength and type of signal produced.

Docking is frequently used to predict the binding orientation of small molecule drug candidates to their protein targets in order to in turn predict the affinity and activity of the small molecule. Hence docking plays an important role in the rational design of drugs. Given the biological and pharmaceutical significance of molecular docking, considerable efforts have been directed towards improving the methods used to predict docking [39].

Molecular docking can be thought of as a problem of *"lock-and-key"*, where one is interested in finding the correct relative orientation of the *"key"* which will open up the *"lock"* (where on the surface of the lock is the key hole, which direction to turn the key after it is inserted, etc.). Here, the protein can be thought of as the "lock" and the ligand can be thought of as a "key". Molecular docking may be defined as an optimization problem, which would describe the "best-fit" orientation of a ligand that binds to a particular protein of interest. However, since both the ligand and the protein are flexible, a*"hand-in-glove"* analogy is more appropriate than *"lock-and-key"*. During the course of the process, the ligand and the protein adjust their conformation to achieve an overall "best-fit" and this kind of conformational adjustments resulting in the overall binding is referred to as "induced-fit".

The focus of molecular docking is to computationally simulate the molecular recognition process. The aim of molecular docking is to achieve an optimized conformation for both the protein and ligand and relative orientation between protein and ligand such that the free energy of the overall system is minimized [40].

2.5.2 Docking approaches

Two approaches are particularly popular within the molecular docking community. One approach uses a matching technique that describes the protein and the ligand as complementary surfaces. The second approach simulates the actual docking process in which the ligand-protein pair wise interaction energies are calculated. Both approaches have significant advantages as well as some limitations. These are outlined below.

The simulation of the docking process as such is a much more complicated process. In this approach, the protein and the ligand are separated by some physical distance, and the ligand finds its position into the protein's active site after a certain number of "moves" in its conformational space. The moves incorporate rigid body transformations such as translations and rotations, as well as internal changes to the ligand's structure including torsion angle rotations [41, 42]. Each of these moves in the conformation space of the ligand induces a total energetic cost of the system, and hence after every move the total energy of the system is calculated. The obvious advantage of the method is that it is more amenable to incorporate ligand flexibility into its modeling whereas shape complementarity techniques have to use some ingenious methods to incorporate flexibility in ligands. Another advantage is that the process is physically closer to what happens in reality, when the protein and ligand approach each other after molecular recognition. A clear disadvantage of this technique is that it takes longer time to evaluate the optimal pose of binding since they have to explore a rather large energy landscape. However grid-based techniques as well as fast optimization methods have significantly ameliorated these problems.

2.5.3 Mechanics of Docking

To perform a docking screen, the first requirement is a structure of the protein of interest. Usually the structure has been determined using a biophysical technique such as x-ray crystallography, or less often, NMR spectroscopy. This protein structure and a database of potential ligands serve as inputs to a docking program. The success of a docking program depends on two components: the search algorithm and the scoring function.

2.5.4 Ligand flexibility

Conformations of the ligand may be generated in the absence of the receptor and subsequently docked or conformations may be generated on-the-fly in the presence of the receptor binding cavity, or with full rotational flexibility of every dihedral angle using fragment based docking. Force field energy evaluation is most often used to select energetically reasonable conformations, but knowledge-based methods have also been used.

2.6 Applications

A binding interaction between a small molecule ligand and an enzyme protein may result in activation or inhibition of the enzyme. If the protein is a receptor, ligand binding may result in agonism or antagonism. Docking is most commonly used in the field of drug design - most drugs are small organic molecules, and docking may be applied to:

- Hit identification – docking combined with a scoring function can be used to quickly screen large databases of potential drugs in silico to identify molecules that are likely to bind to protein target of interest (see virtual screening).
- Lead optimization – docking can be used to predict in where and in which relative orientation a ligand binds to a protein (also referred to as the binding mode or pose). This information may in turn be used to design more potent and selective analogs.
- Bioremediation – Protein ligand docking can also be used to predict pollutants that can be degraded by enzymes.

Chapter-3 Materials and Methods

3.1 Databases:

3.1.1. Swiss-Prot:

Swiss-Prot is a manually curated biological database of protein sequences. Swiss-Prot was created in 1986 by Amos Bairoch during his PhD and developed by the Swiss Institute of Bioinformatics and the European Bioinformatics Institute. Swiss-Prot strives to provide reliable protein sequences associated with a high level of annotation (such as the description of the function of a protein, its domains structure, post-translational modifications, variants, etc.), a minimal level of redundancy and high level of integration with other databases.

3.1.2 Protein Data Bank

The Protein Data Bank (PDB) is a repository for the 3-D structural data of large biological molecules, such as proteins and nucleic acids. The PDB originated as a grassroots effort. In 1971, Walter Hamilton of the Brookhaven National Laboratory agreed to set up the data bank at Brookhaven. Upon Hamilton's death in 1973, Tom Koeztle took over direction of the PDB. In January 1994, Joel Sussman was appointed head of the PDB. In October 1998, the PDB was transferred to the Research Collaboratory for Structural Bioinformatics (RCSB); the transfer was completed in June 1999. The new director was Helen M. Berman of Rutgers University (one of the member institutions of the RCSB). In 2003, with the formation of the wwPDB, the PDB became an international organization. The PDB is overseen by an organization called the Worldwide Protein Data Bank. The PDB is a key resource in areas of structural biology, such as structural genomics. Most major scientific journals, and some funding agencies, such as the NIH in the USA, now require scientists to submit their structure data to the PDB. If the contents of the PDB are thought of as primary data, then there are hundreds of derived (i.e., secondary) databases that categorize the data differently. For example, both SCOP and CATH categorize structures according to type of structure and assumed evolutionary relations; GO categorize structures based on genes.

3.1.3 PubChem

PubChem is a database of chemical molecules. The system is maintained by the National Center for Biotechnology Information (NCBI), a component of the National Library of Medicine, which is part of the United States National Institutes of Health (NIH). PubChem can be accessed for free through a web user interface. Millions of compound structures and descriptive datasets can be freely downloaded via FTP. PubChem contains substance descriptions and small molecules with fewer than 1000 atoms and 1000 bonds. The American Chemical Society tried to get the U.S. Congress to restrict the operation of PubChem, because they claim it competes with their Chemical Abstracts Service.. More than 80 database vendors contribute to the growing PubChem database.

3.1.4 Drug Bank

The Drug Bank database available at the University of Alberta is a bioinformatics and cheminformatics resource that combines detailed drug (i.e. chemical, pharmacological and pharmaceutical) data with comprehensive drug target (i.e. sequence, structure, pathway) information. The database contains nearly 4800 drug entries including:

> 1480 FDA-approved small molecule drugs,

128 FDA-approved biotech (protein/peptide) drugs,

> 71 nutraceuticals, and

> 3200 experimental drugs.

More than 2500 protein (i.e. drug target, non-redundant) sequences are linked to these drug entries. Each Drug Card entry contains more than 100 data fields with half of the information being devoted to drug/chemical data and the other half devoted to drug target or protein data.

It is maintained by David Wishart and Craig Knox

3.2 Servers

3.2.1 SBASE:

SBASE is a collection of protein domain sequences designed to facilitate the detection of distant similarities typically found between modules of multidomain proteins. A multidomain protein can share a biologically significant sequence pattern with a number of different, functionally related proteins or protein domains, even though the sequence alignments may not be highly significant in the mathematical sense. SBASE can be considered as a conversion of

the protein sequence database into a format that facilitates detection of such functional and structural similarities. The current Release 4.0 of SBASE contains over 60 000 annotated protein sequence segments consistently named by structure, function, biased composition, binding specificity and/or similarity to other proteins. The format of the database is such that it can be searched with standard programs, like FASTA or BLAST3, and the information given allows the prediction of function and the direct detection of potential domain homologies.

3.2.2.BLAST

BLAST is actually a family of programs (all included in the blastall executable). The following are some of the programs Blast.

BLAST searches for high scoring sequence alignments between the query sequence and sequences in the database using a heuristic approach that approximates the Smith-Waterman algorithm. The exhaustive Smith-Waterman approach is too slow for searching large genomic databases such as GenBank. Therefore, the BLAST algorithm uses a heuristic approach that is slightly less accurate than Smith-Waterman but over 50 times faster. The speed and relatively good accuracy of BLAST are the key technical innovation of the BLAST programs and arguably why the tool is the most popular bioinformatics search tool, ranked mostly in order of importance:

a. Nucleotide-nucleotide BLAST (blastn): This program, given a DNA query, returns the most similar DNA sequences from the DNA database that the user specifies.

b. Protein-protein BLAST (blastp): This program, given a protein query, returns the most similar protein sequences from the protein database that the user specifies.

c. Position-Specific Iterative BLAST (PSI-BLAST): One of the more recent BLAST programs, this program is used for finding distant relatives of a protein. First, a list of all closely related proteins is created. Then these proteins are combined into a "profile" that is a sort of average sequence. A query against the protein database is then run using this profile, and a larger group of proteins found. This larger group is used to construct another profile, and the process is repeated. By including related proteins in the search PSI-BLAST is much more sensitive in picking up distant evolutionary relationships than the standard protein-protein BLAST.

d. Nucleotide 6-frame translation-protein (blastx): This program compares the six-frame conceptual translation products of a sequence database.

e. Nucleotide 6-frame translation-nucleotide 6-frame translation (tblastx): This program is the slowest of the BLAST family. It translates the query nucleotide sequence in all six possible frames and compares it against the six-frame translations of a nucleotide sequence database. The purpose of tblastx is to find very distant relationships between nucleotide sequences.

f. Protein-nucleotide 6-frame translation (tblastn): This program compares a protein query against the six-frame translations of a nucleotide sequence database.

g. Large numbers of query sequences (megablast): When comparing large numbers of input sequences via the command-line BLAST, "megablast" is much faster than running BLAST multiple times. It basically concatenates many input sequences together to form a large sequence before searching the BLAST database, then post-analyze the search results to glean individual alignments and statistical values" nucleotide query sequence (both strands) against a protein.

3.3 Alignment

Up to five template structures per batch are superposed using an iterative least squares algorithm. A structural alignment is generated after removing incompatible templates, i.e. omitting structures with high $C\alpha$ root mean square deviations to the first template. A local pair-wise alignment of the target sequence to the main template structures is calculated, followed by a heuristic step to improve the alignment for modeling purposes. The placement of insertions and deletions is optimized considering the template structure context. In particular, isolated residues in the alignment (‘islands’) are moved to the flanks to facilitate the loop building process.

3.4 Softwares

3.4.1 MOLSOFT

Molsoft is a La Jolla, California based company that is a primary source of new breakthrough technologies in: molecular graphics and visualization, molecular modeling, docking and virtual screening, computational biology and cheminformatics. Icm program is based on a stochastic algorithm that relies on global optimization of the entire flexible ligand in the receptor field. Building and validating structural models of protein targets.

Identifying biological ligand binding sites or new sites for allosteric regulation of a protein of interest.

Evaluating and ranking drug targets, including protein-protein interaction interfaces, designing strategies for rational drug design.

3.4.2 CASTp

Computed Atlas of Surface Topography of proteins (CASTp) provides an online resource for locating, delineating and measuring concave surface regions on three-dimensional structures of proteins. These include pockets located on protein surfaces and voids buried in the interior of proteins. The measurement includes the area and volume of pocket or void by solvent accessible surface model (Richards' surface) and by molecular surface model (Connolly's surface), all calculated analytically. CASTp can be used to study surface features and functional regions of proteins. CASTp includes a graphical user interface, flexible interactive visualization, as well as on-the-fly calculation for user uploaded structures.

3.4.3 Chemsketch

ACD/ChemSketch is an integrated software package from Advanced Chemistry Development Inc. for drawing chemical structures, reactions, schematic diagrams and designing other chemistry-related reports and presentations. Structure mode for drawing chemical structures and calculating their properties.

3.5 Docking

Docking studies are computational techniques for the exploration of the possible binding modes of a substrate to a given receptor, enzyme or other binding site. Docking studies have become nearly indispensable for study of macromolecular structures and interactions. Mechanical model construction requires heroic patience and endurance to complete a structure which may contain several thousand atoms while computer graphics can build and display in seconds. Macromolecular modeling by Docking studies provides most detailed possible view of drug-receptor interaction and has created a new rational approach to drug design where the structure of drug is designed based on its fit to three dimensional structures of receptor site, rather than by analogy to other active structures of random leads.

3.6 Methodology

3.6.1 3D model building

Homology modeling, also known as comparative modeling refers to constructing an atomic-resolution model of the "*target*" protein from its amino acid sequence and an experimental three-dimensional structure of a related homologous protein (the "*template*"). Homology modeling relies on the identification of one or more known protein structures likely to resemble the structure of the query sequence, and on the production of an alignment that maps residues in the query sequence to residues in the template sequence. The sequence alignment and template structure are then used to produce a structural model of the target. Because protein structures are more conserved than DNA sequences, detectable levels of sequence similarity usually imply significant structural similarity.

Suppose you want to know the 3D structure of a target protein that has not been solved empirically by X-ray crystallography or NMR. You have only the sequence. If an empirically determined 3D structure is available for a sufficiently similar protein (50% or better sequence identity would be good), you can use software that arranges the backbone of your sequence identically to this template. This is called "homology modeling". It is, at best, moderately accurate for the positions of alpha carbons in the 3D structure, in regions where the sequence identity is high. It is inaccurate for the details of side chain positions, and for inserted loops with no matching sequence in the solved structure.

The optimization procedure is a variable target function method that applies the conjugate gradients algorithm to positions of all non-hydrogen atoms. The query sequence from Homo sapiens was submitted to domain fishing server Signal transducer and activator of transcription 4 prediction. The predicted domain was searched to find out the related protein structure to be used as a template by the BLAST (Basic Local Alignment Search Tool) program against PDB (Protein Databank). Sequence that showed maximum identity with high score and less e-value was aligned and Bile salt export pump was used as a reference structure to build a 3D model for. The sequence of Bile salt export pump was obtained from UNIPROT.

The co-ordinates for the structurally conserved regions (SCRs) for Signal transducer and activator of transcription 4 were assigned from the template using multiple sequence alignment, based on the Needleman-Wunsch algorithm. The structure having the least modeller objective function, obtained from the modeller was improved by molecular dynamics and equilibration

methods using NAMD 2.5 software using CHARMM27 force field for lipids and proteins along with the TIP3P model for water. The energy of the structure was minimized with 1, 00, 00 steps. A cutoff of 12 Å (switching function starting at 10 Å) for van der Waals interactions was assumed. No periodic boundary conditions were included in this study. An integration time step of 2 fs was used, permitting a multiple time-stepping algorithm to be employed in which interactions involving covalent bonds were computed every time step, short-range non bonded interactions were computed every two time steps and long-range electrostatic forces were computed every four time steps.

The pair list of the non bonded interaction was recalculated every ten time steps with a pair list distance of 13.5 Å. The short-range non bonded interactions were defined as van der Waals and electrostatics interactions between particles within 12 Å. A smoothing function was employed for the van der Waals interactions at a distance of 10 Å. CHARMM27 [force-field parameters were used in all simulations in this study. The equilibrated system was simulated for 1 ps with a 500 kcal/mol/Å2 restraint on the protein backbone under 1 atm constant pressure and 310 K constant temperature (NPT) and the Langevin damping coefficient was set to 5 ps unless otherwise stated. Finally, the structure having the least energy with low RMSD (Root Mean Square Deviation) was used for further studies. In this step, the quality of the initial model was improved. The final structure obtained was analyzed by Ramachandran's map using PROCHECK (Programs to check the Stereo chemical Quality of Protein Structures) and environment profile using ERRAT graph (Structure Evaluation server). This model was used for the identification of active site and for docking of the substrate with the enzyme.

3.6.2 Active site Identification

Active site of Signal transducer and activator of transcription 4 was identified using CASTp server. A new program, CASTp, for automatically locating and measuring protein pockets and cavities, is based on precise computational geometry methods, including alpha shape and discrete flow theory. CASTp identifies and measures pockets and pocket mouth openings, as well as cavities. The program specifies the atoms lining pockets, pocket openings, and buried cavities; the volume and area of pockets and cavities; and the area and circumference of mouth openings.

3.6.3 Docking method

Docking was carried out using GOLD (Genetic Optimization of Ligand Docking) software which is based on genetic algorithm (GA). This method allows as partial flexibility of protein and full flexibility of ligand. The compounds are docked to the active site of the Bcl2L10. The interaction of these compounds with the active site residues are thoroughly studied using molecular mechanics calculations. The parameters used for GA were population size (100), selection pressure (1.1), number of operations (10,000), number of island (1) and niche size (2). Operator parameters for crossover, mutation and migration were set to 100, 100 and 10 respectively. Default cutoff values of 3.0 A° (dH-X) for hydrogen bonds and 6.0 A° for vanderwaals were employed. During docking, the default algorithm speed was selected and the ligand binding site in the alpha glucosidase was defined within a 10 A° radius with the centroid as CE atom of ALA410. The number of poses for each inhibitor was set 100, and early termination was allowed if the top three bound conformations of a ligand were within 1.5A° RMSD. After docking, the individual binding poses of each ligand were observed and their interactions with the protein were studied. The best and most energetically favorable conformation of each ligand was selected.

3.6.4 Gold Score fitness function

Gold Score performs a force field based scoring function and is made up of four components: 1. Protein-ligand hydrogen bond energy (external H-bond); 2. Protein-ligand vander Waals energy (external vdw); 3. Ligand internal vander Waals energy (internal vdw); 4. Ligand intramolecular hydrogen bond energy (internal- H- bond). The external vdw score is multiplied by a factor of 1.375 when the total fitness score is computed. This is an empirical correction to encourage protein-ligand hydrophobic contact. The fitness function has been optimized for the prediction of ligand binding positions.

GoldScore = S (hb_ext) + S (vdw_ext) + S (hb_int) + S (vdw_int)

Where S (hb_ext) is the protein-ligand hydrogen bond score, S (vdw_ext) is the protein-ligand van der Waals score, S (hb_int) is the score from intramolecular hydrogen bond in the ligand and S (vdw_int) is the score from intramolecular strain in the ligand.

Chapter-4 Results

4.1 Homology Modeling of Bcl2L10Protein

A high level of sequence identity should guarantee more accurate alignment between the target sequence and template structure. In the results of BLAST search against PDB, 2KUA (chain A) has a high level of sequence identity and the identity of the reference protein with the Bcl2L10 are 48%. Structurally conserved regions (SCRs) for the model and the template were determined by superimposition of the two structures and multiple sequence alignment.

```
template        MADSQDPLHERTRRLLSDYIFFCAREPDTPEPPPTSVEAALLRSVTRQIQQEHQEFFSSF
query           ---MADPLRERTELLLADYLGYCAREPGTPEPAPSTPEAAVLRSAAARLRQIHRSFFSAY
                   ***:***. **:**: :*****.****.*:: ***:***.: :::* *:.***::

template        CESRGNRLELVKQMADKLLSKDQDFSWSQLVMLLAFAGTLMNQGPYMAVKQKRD------
query           LGYPGNRFELVALMADSVLSDSPGPTWGRVVTLVTFAGTLLERGPLVTARWKKWGFQPRL
                   ***:***  ***.:**.. . :*.::* *::*****:::** ::.: *:

template        LGNRVIVTRDCCLIVNFLYNLLMGRRHRARLEALGGWDGFCRFFKNPLPLG---------
query           KEQEGDVARDCQRLVALLSSRLMG-QHRAWLQAQGGWDGFCHFFRTPFPLAFWRKQLVQA
                  :.  *:***  :* :* . *** :*** *:* *******:**:.*:**.

template        ------------------
query           FLSCLLTTAFIYLWTRLL
```

Figure 9: Clustal multiple sequence alignment

In the following study, we have chosen 2KUA as a reference structure for modeling domain. Coordinates from the reference protein 2KUA to the SCRs, structurally variable regions (SVRs), N-termini and C-termini were assigned to the target sequence based on the satisfaction of spatial restraints. In the modeller we will get a 20 PDB out of which we select a least energy. The energy unit will be in kilo joule. All side chains of the model protein were set by rotamers. The final stable structure of the BCL2L10 protein obtained is shown in Figure 2. By the help of SPDBV it is evident that Bile salt export pump domain has 8 helices.

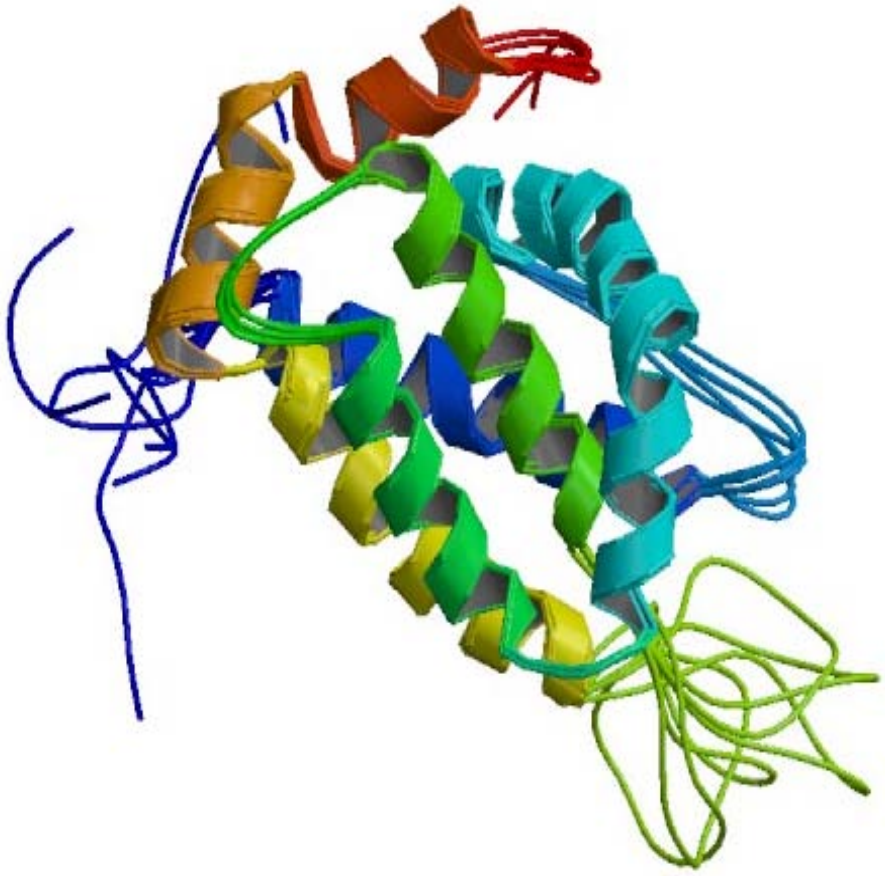


Figure 10: Structure of the BCL2L10 protein

The final structure was further checked by verify3D graph and the results have been shown in Figure 12: The overall scores indicates acceptable protein environment.

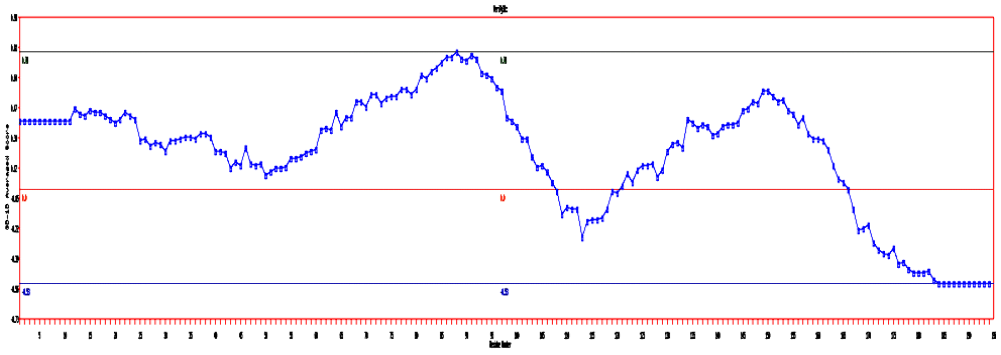


Figure 11: Graphical representation for verify 3D

4.2 Validation of BCL2L10

After the refinement process, validation of the model was carried out using Ramachandran plot calculations computed with the PROCHECK program. The distributions of the Ramachandran plots of non-glycine, non-proline residues are summarized in Table 1. The RMSD (Root Mean Square deviation) deviation for covalent bonds and covalent angles relative to the standard dictionary of BCL2L10 was -5.27 and -0.55 Å. Altogether 99.4 % of the residues of BCL2L10 was in favored and allowed regions. The overall PROCHECK G-factor of BCL2L10 was – 2.32 and verify3D environment profile was good.

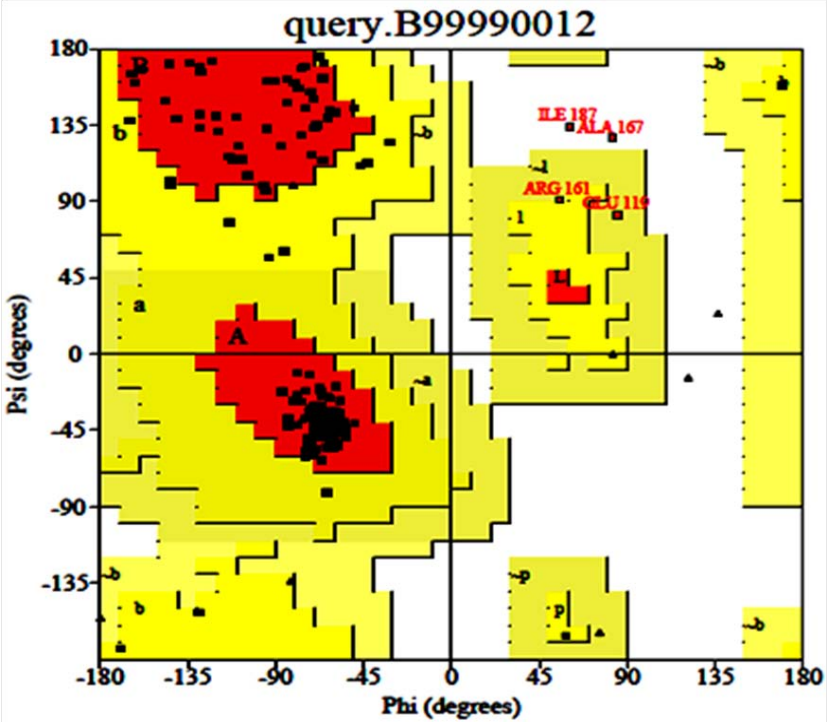


Figure 12: Ramachandran Plot using RAMPAGE server

Table 1: % of residue falling in the core region of the Ramachandran's plot

% of residue in most favored regions	87.3
% of residue in the additionally allowed zones	10.3
% of residue in the generously regions	1.2

% of residue in disallowed regions	1.2
% of non-glycine and non-proline residues	100.0

4.3 Superimposition of 2KUA with BCL2L10 domain

The structural superimposition of 2KUA template and BCL2L10 is shown in Figure 14. The weighted root mean square deviation of trace between the template and final refined models 0.82A^{o}. This final refined model was used for the identification of active site and for docking of the substrate with the domain BCL2L10.

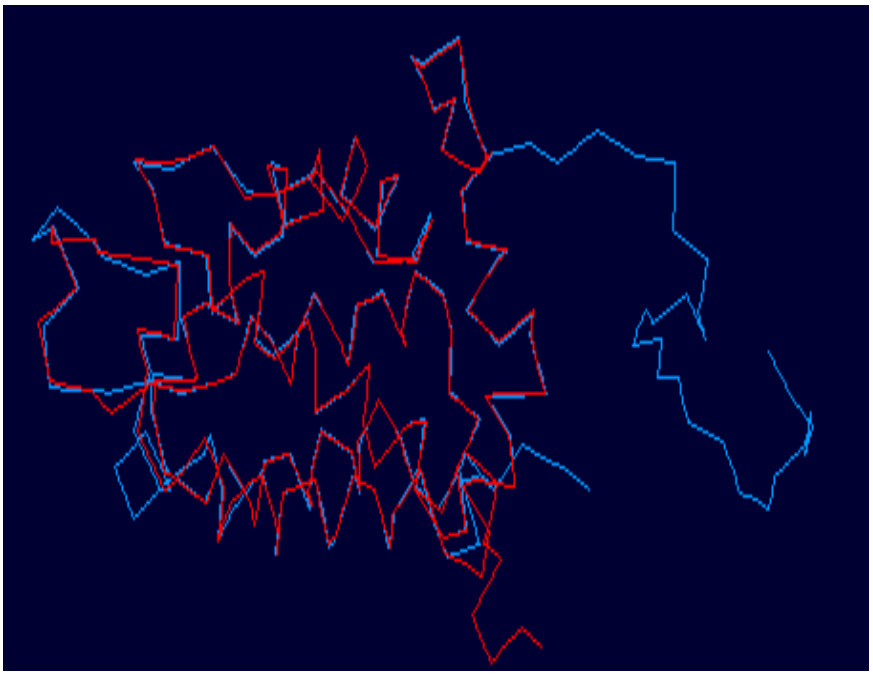


Figure 13: Compare the structure of Bcl2L10(represented in Red color) with 2KUA (represented with Blue color)

4.4 Active site Identification of BCL2L10

After the final model was built, the possible binding sites of BCL2L10 was searched based on the structural comparison of template and the model build and also with CASTP server and was shown in Figure 19. Since, BCL2L10 from Human and the 2KUA are well conserved in both sequence and structure; their biological function should be identical. Infact from the structure-structure comparison of template, final refined model of BCL2L10 domain using

SPDBV program. It was found that secondary structures are highly conserved and the residues, PHE5, LYS7, CYS9, TYR10GLY11,TYR12, CYS13, ILE14, PHE28, LEU30, TYR31, ALA61, VAL62, THR63, PHE82, VAL230, LEU281.

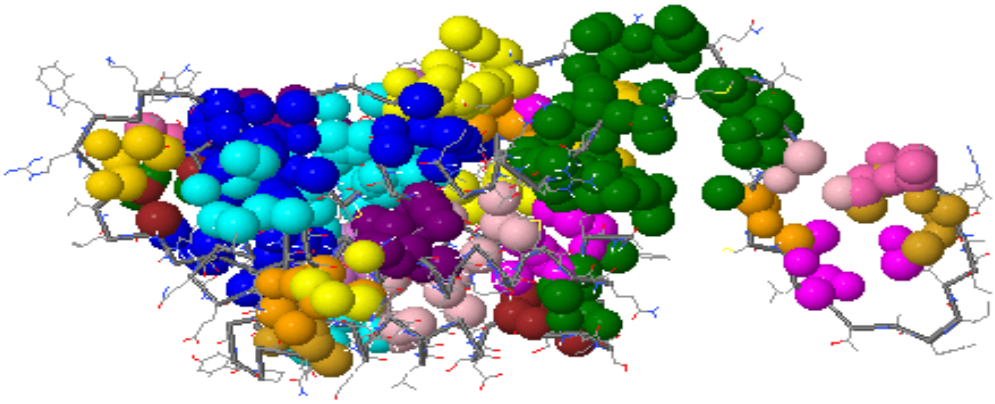
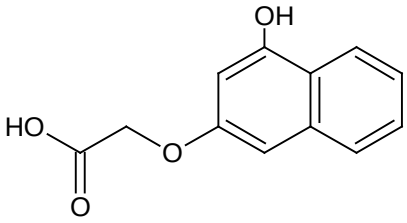


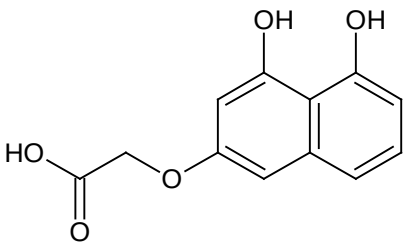
Figure 14: Active site of protein Bcl2L10

Chemical structure (inhibitor) of drugs that designed for using on docking in the study

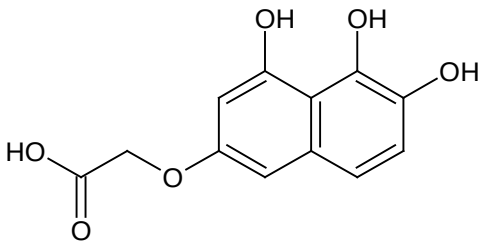
Different compounds were checked for the docking with protein Bcl2L10 as following:

1- 

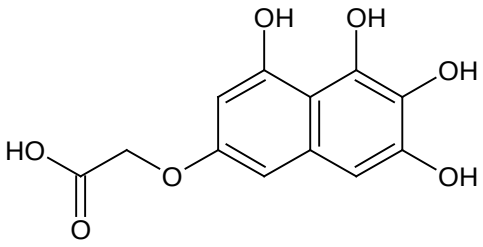
[(4-hydroxy-2-naphthyl)oxy]acetic acid

2- 

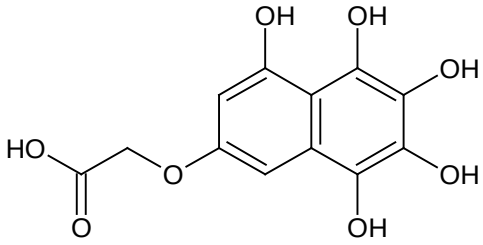
[(4,5-dihydroxy-2-naphthyl)oxy]acetic acid



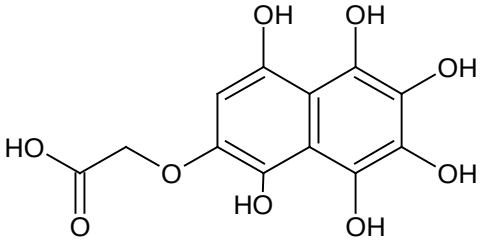
3- [(4,5,6-trihydroxy-2-naphthyl)oxy]acetic acid



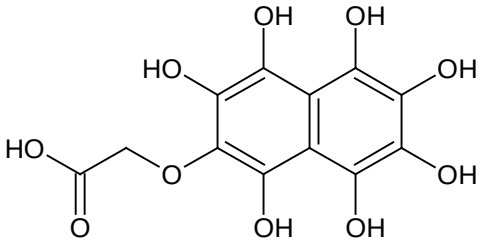
4- [(4,5,6,7-tetrahydroxy-2-naphthyl)oxy]acetic acid



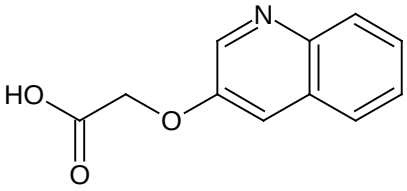
5- [(4,5,6,7,8-pentahydroxy-2-naphthyl)oxy]acetic acid



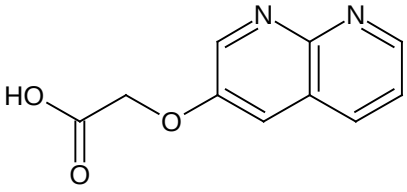
6- [(1,4,5,6,7,8-hexahydroxy-2-naphthyl)oxy]acetic acid

7- 

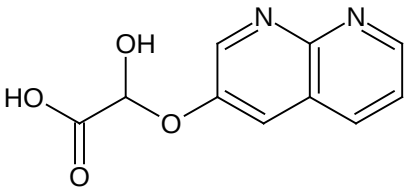
[(1,3,4,5,6,7,8-heptahydroxy-2-naphthyl)oxy]acetic acid

8- 

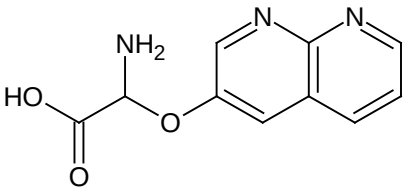
(quinolin-3-yloxy)acetic acid

9- 

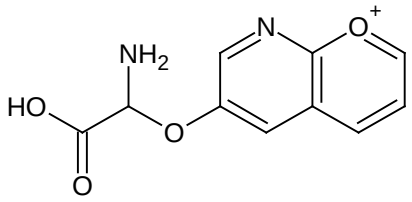
(1,8-naphthyridin-3-yloxy)acetic acid

10- 

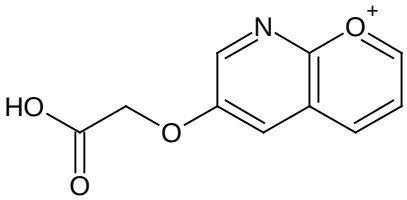
hydroxy(1,8-naphthyridin-3-yloxy)acetic acid

11- 

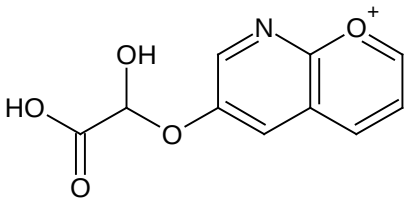
amino(1,8-naphthyridin-3-yloxy)acetic acid



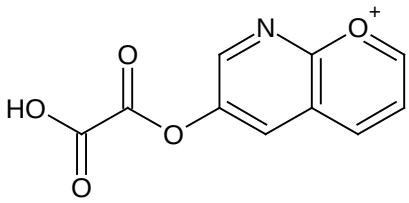
12- 6-[amino(carboxy)methoxy]pyrano[2,3-*b*]pyridin-1-ium



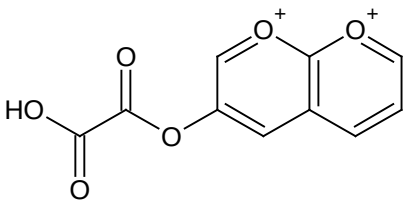
13- 6-(carboxymethoxy)pyrano[2,3-*b*]pyridin-1-ium



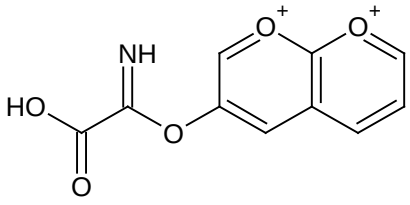
14- 6-[carboxy(hydroxy)methoxy]pyrano[2,3-*b*]pyridin-1-ium



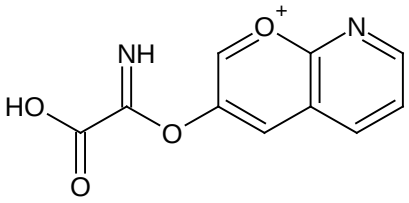
15- 6-[(carboxycarbonyl)oxy]pyrano[2,3-*b*]pyridin-1-ium



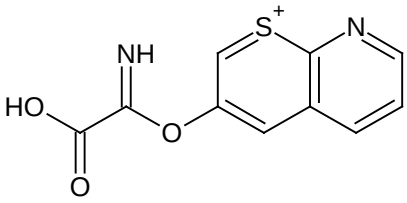
16- 3-[(carboxycarbonyl)oxy]pyrano[2,3-*b*]pyrandiium



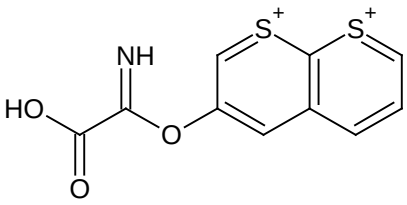
17- 3-[carboxy(imino)methoxy]pyrano[2,3-*b*]pyrandiium



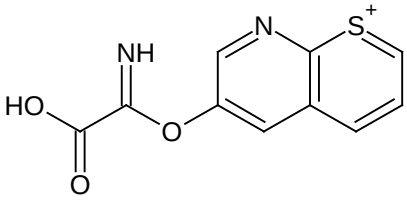
18- 3-[carboxy(imino)methoxy]pyrano[2,3-*b*]pyridin-1-ium



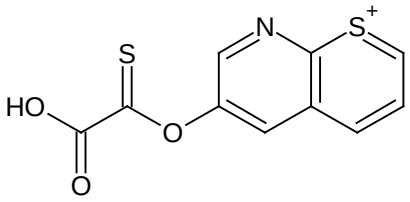
19- 3-[carboxy(imino)methoxy]thiopyrano[2,3-*b*]pyridin-1-ium



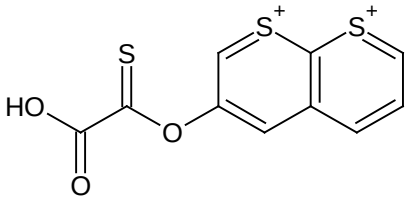
20- 3-[carboxy(imino)methoxy]thiopyrano[2,3-*b*]thiopyrandiium



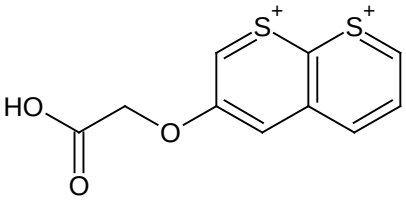
21- 6-[carboxy(imino)methoxy]thiopyrano[2,3-*b*]pyridin-1-ium



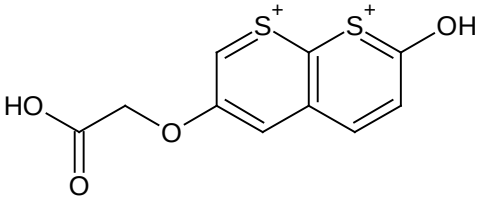
22- 6-[(carboxycarbonothioyl)oxy]thiopyrano[2,3-*b*]pyridin-1-ium



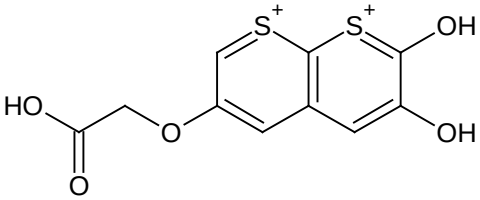
23- 3-[(carboxycarbonothioyl)oxy]thiopyrano[2,3-*b*]thiopyrandiium



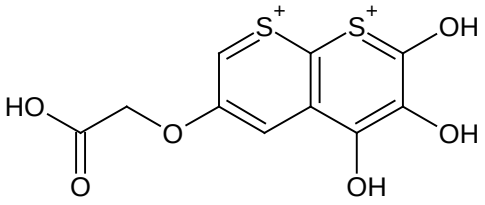
24- 3-(carboxymethoxy)thiopyrano[2,3-*b*]thiopyrandiium



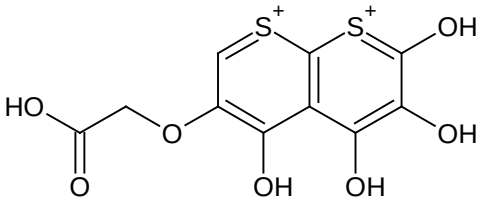
25- 6-(carboxymethoxy)-2-hydroxythiopyrano[2,3-*b*]thiopyrandiium



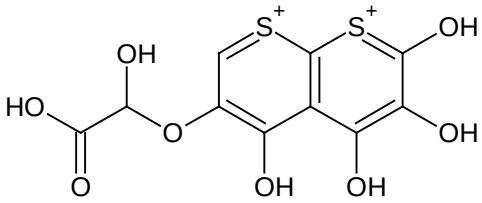
26- 6-(carboxymethoxy)-2,3-dihydroxythiopyrano[2,3-*b*]thiopyrandiium



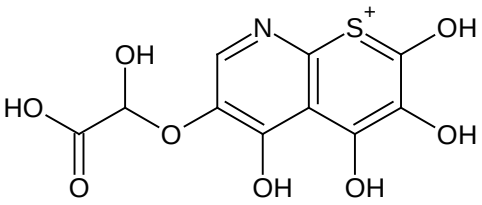
27- 6-(carboxymethoxy)-2,3,4-trihydroxythiopyrano[2,3-*b*]thiopyrandiium



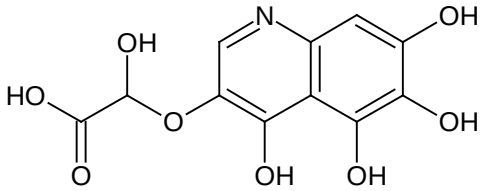
28- 6-(carboxymethoxy)-2,3,4,5-tetrahydroxythiopyrano[2,3-*b*]thiopyrandiium



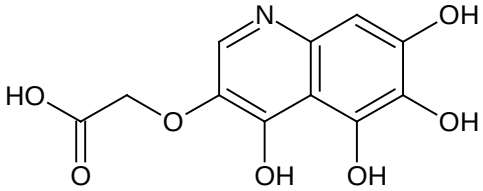
29- 6-[carboxy(hydroxy)methoxy]-2,3,4,5-tetrahydroxythiopyrano[2,3-*b*]thiopyrandiium



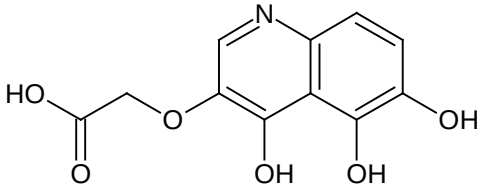
30- 6-[carboxy(hydroxy)methoxy]-2,3,4,5-tetrahydroxythiopyrano[2,3-*b*]pyridin-1-ium



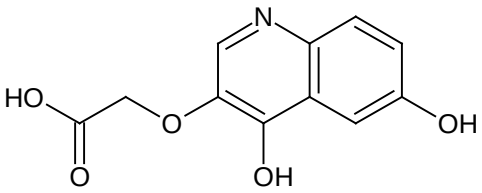
31- hydroxy[(4,5,6,7-tetrahydroxyquinolin-3-yl)oxy]acetic acid



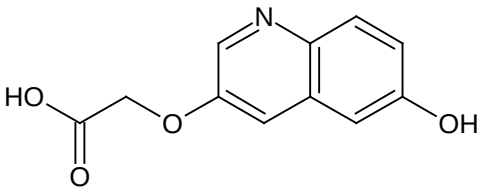
32- [(4,5,6,7-tetrahydroxyquinolin-3-yl)oxy]acetic acid



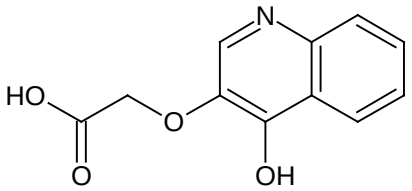
33- [(4,5,6-trihydroxyquinolin-3-yl)oxy]acetic acid



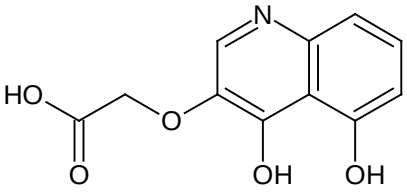
34- [(4,6-dihydroxyquinolin-3-yl)oxy]acetic acid



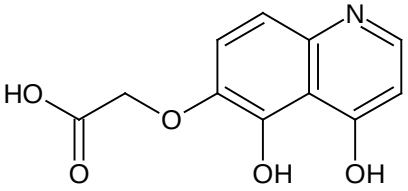
35- [(6-hydroxyquinolin-3-yl)oxy]acetic acid



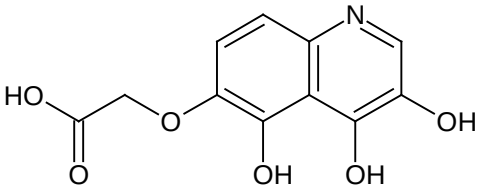
36- [(4-hydroxyquinolin-3-yl)oxy]acetic acid



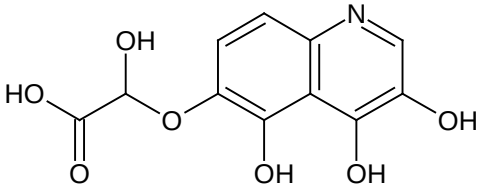
37- [(4,5-dihydroxyquinolin-3-yl)oxy]acetic acid



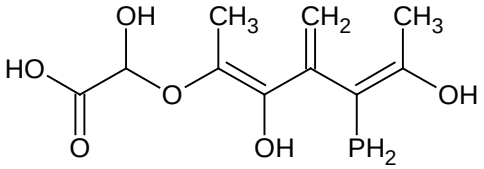
38- [(4,5-dihydroxyquinolin-6-yl)oxy]acetic acid



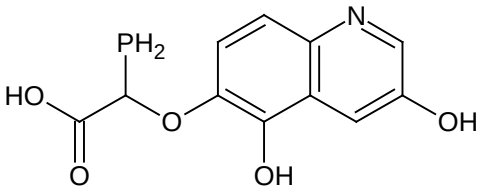
39- [(3,4,5-trihydroxyquinolin-6-yl)oxy]acetic acid



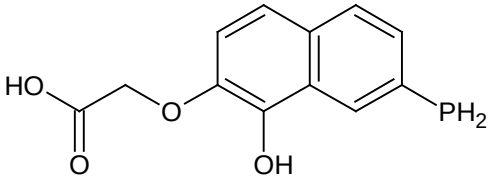
40- hydroxy[(3,4,5-trihydroxyquinolin-6-yl)oxy]acetic acid



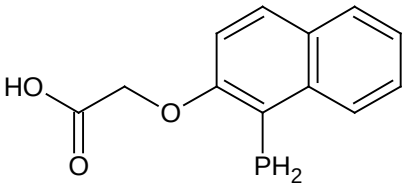
41- [(3,5-dihydroxy-4-phosphinoquinolin-6-yl)oxy](hydroxy)acetic acid



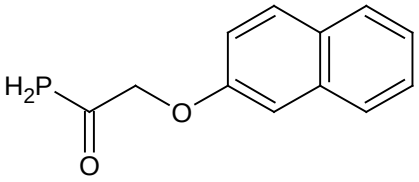
42- [(3,5-dihydroxyquinolin-6-yl)oxy](phosphino)acetic acid



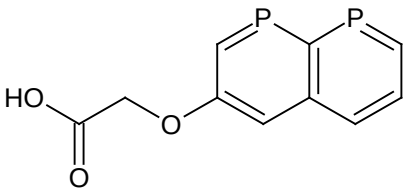
43- [(1-hydroxy-7-phosphino-2-naphthyl)oxy]acetic acid



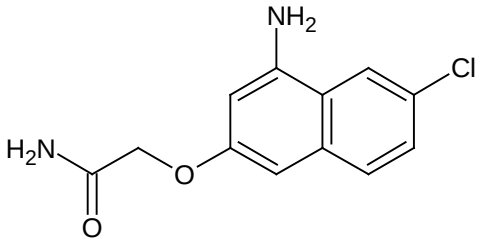
44- [(1-phosphino-2-naphthyl)oxy]acetic acid



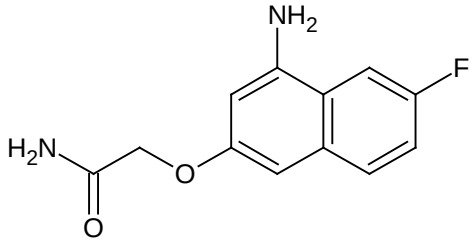
45- [(2-naphthyloxy)acetyl]phosphine



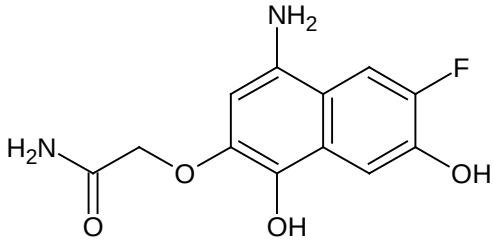
46- (phosphinino[2,3-*b*]phosphinin-3-yloxy)acetic acid



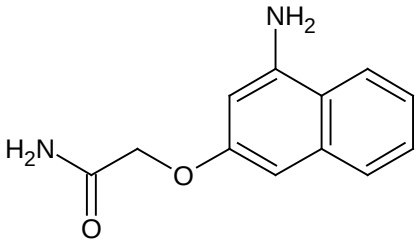
47- 2-[(4-amino-6-chloro-2-naphthyl)oxy]acetamide



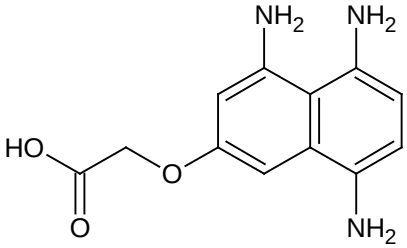
48- 2-[(4-amino-6-fluoro-2-naphthyl)oxy]acetamide



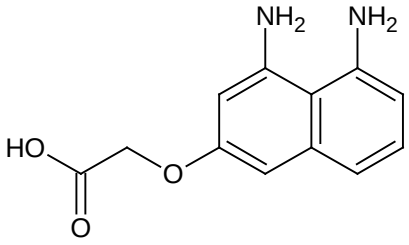
49- 2-[(4-amino-6-fluoro-1,7-dihydroxy-2-naphthyl)oxy]acetamide



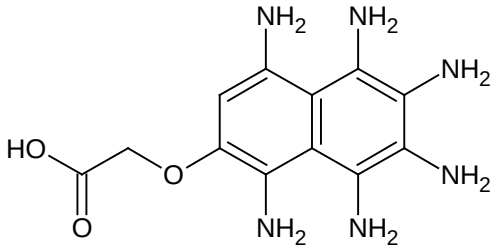
50- 2-[(4-amino-2-naphthyl)oxy]acetamide



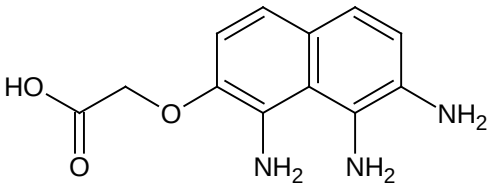
51- [(4,5,8-triamino-2-naphthyl)oxy]acetic acid



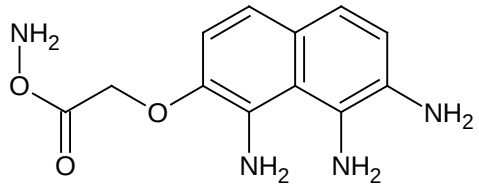
52- [(4,5-diamino-2-naphthyl)oxy]acetic acid



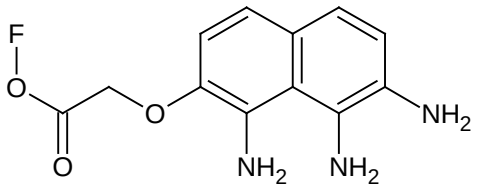
53- [(1,4,5,6,7,8-hexaamino-2-naphthyl)oxy]acetic acid



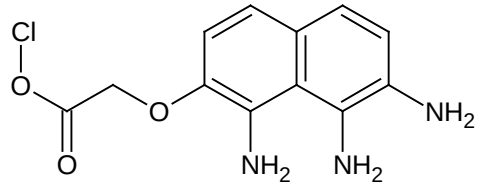
54- [(1,7,8-triamino-2-naphthyl)oxy]acetic acid



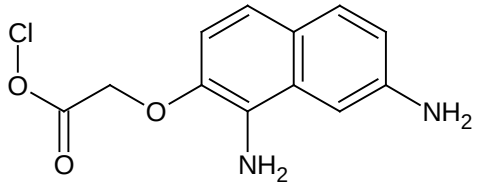
55- 7-[2-(aminooxy)-2-oxoethoxy]naphthalene-1,2,8-triamine



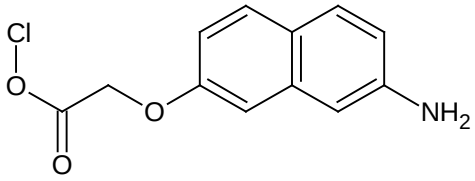
56- 7-[2-(fluorooxy)-2-oxoethoxy]naphthalene-1,2,8-triamine



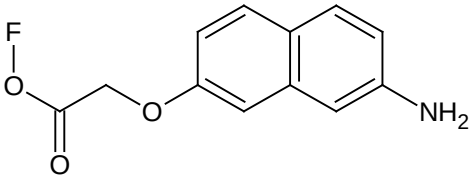
57- 7-[2-(chlorooxy)-2-oxoethoxy]naphthalene-1,2,8-triamine



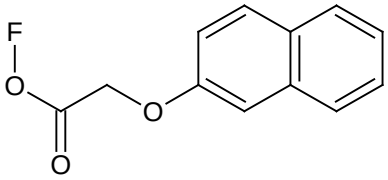
58- 2-[2-(chlorooxy)-2-oxoethoxy]naphthalene-1,7-diamine



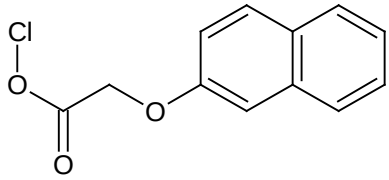
59- 7-[2-(chlorooxy)-2-oxoethoxy]naphthalen-2-amine



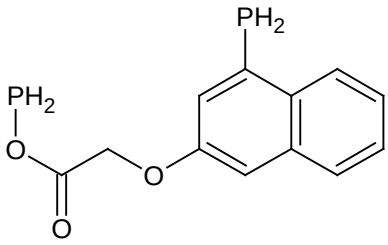
60- 7-[2-(fluorooxy)-2-oxoethoxy]naphthalen-2-amine



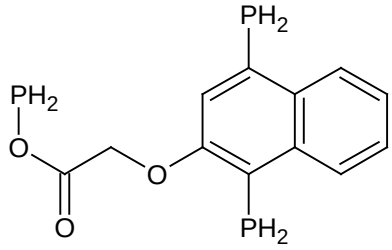
61- 2-[2-(fluorooxy)-2-oxoethoxy]naphthalene



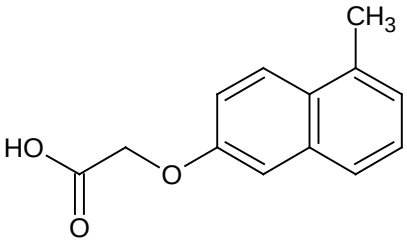
62- 2-[2-(chlorooxy)-2-oxoethoxy]naphthalene



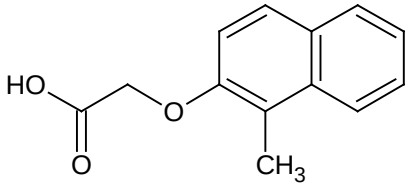
63- {3-[2-oxo-2-(phosphinooxy)ethoxy]-1-naphthyl}phosphine



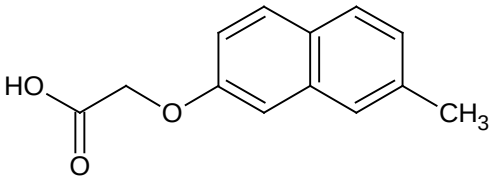
64- {2-[2-oxo-2-(phosphinooxy)ethoxy]naphthalene-1,4-diyl}diphosphine



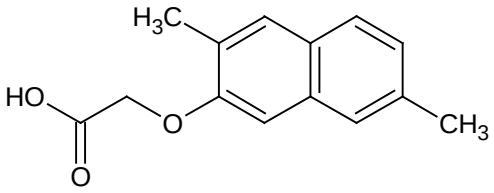
65- [(5-methyl-2-naphthyl)oxy]acetic acid

66- 

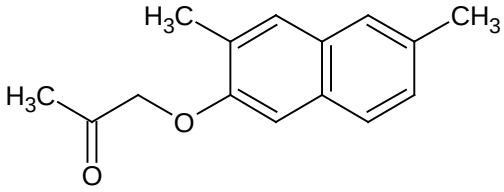
[(1-methyl-2-naphthyl)oxy]acetic acid

67- 

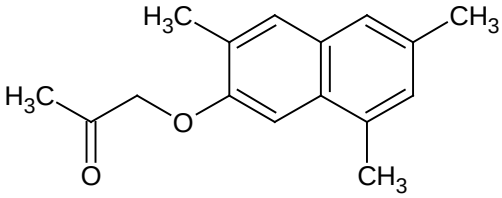
[(7-methyl-2-naphthyl)oxy]acetic acid

68- 

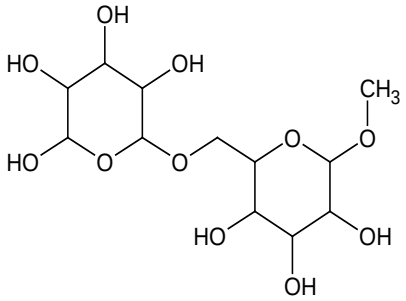
[(3,7-dimethyl-2-naphthyl)oxy]acetic acid

69- 

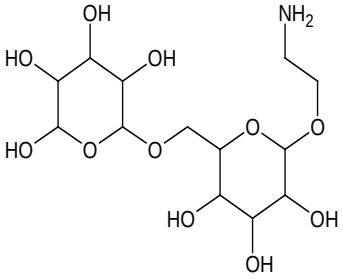
1-[(3,6-dimethyl-2-naphthyl)oxy]acetone

70- 

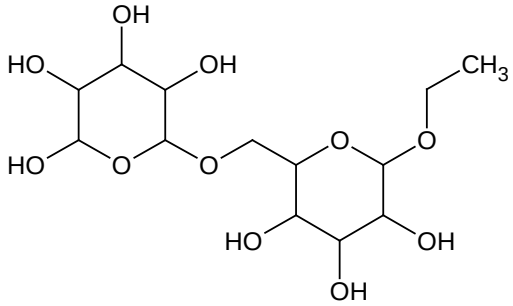
1-[(3,6,8-trimethyl-2-naphthyl)oxy]acetone



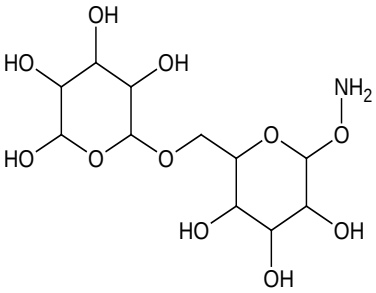
71- 6-[(3,4,5-trihydroxy-6-methoxytetrahydro-2*H*-pyran-2-yl)methoxy]tetrahydro-2*H*-pyran-2,3,4,5-tetrol



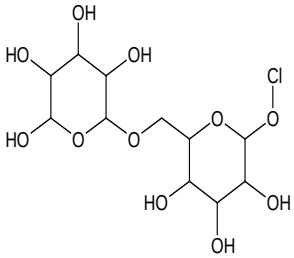
72- 6-{[6-(2-aminoethoxy)-3,4,5-trihydroxytetrahydro-2*H*-pyran-2-yl]methoxy}tetrahydro-2*H*-pyran-2,3,4,5-tetrol



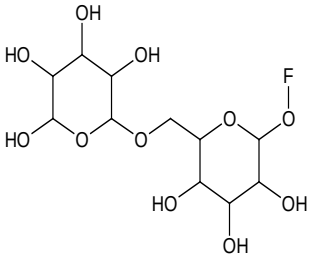
73- 6-[(6-ethoxy-3,4,5-trihydroxytetrahydro-2*H*-pyran-2-yl)methoxy]tetrahydro-2*H*-pyran-2,3,4,5-tetrol



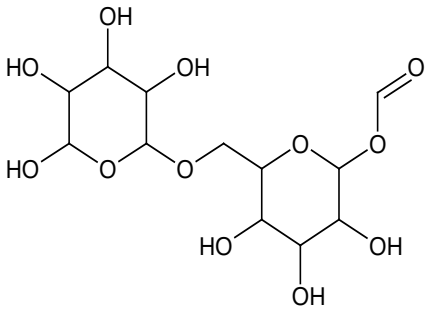
74- 6-{[6-(aminooxy)-3,4,5-trihydroxytetrahydro-2*H*-pyran-2-yl]methoxy}tetrahydro-2*H*-pyran-2,3,4,5-tetrol



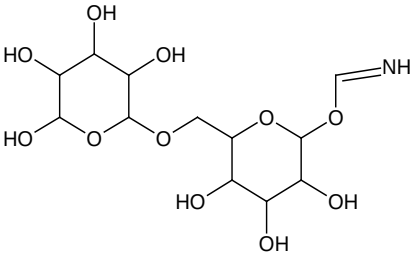
75- 3,4,5-trihydroxy-6-{[(3,4,5,6-tetrahydroxytetrahydro-2*H*-pyran-2-yl)oxy]methyl}tetrahydro-2*H*-pyran-2-yl hypochlorite



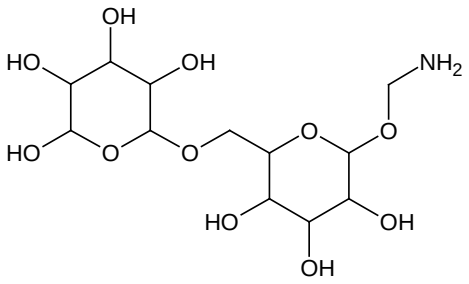
76- 3,4,5-trihydroxy-6-{[(3,4,5,6-tetrahydroxytetrahydro-2*H*-pyran-2-yl)oxy]methyl}tetrahydro-2*H*-pyran-2-yl hypofluorite



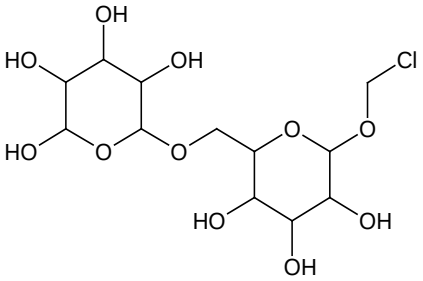
77- 3,4,5-trihydroxy-6-{[(3,4,5,6-tetrahydroxytetrahydro-2*H*-pyran-2-yl)oxy]methyl}tetrahydro-2*H*-pyran-2-yl formate



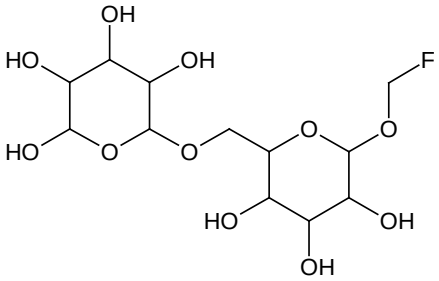
78- 3,4,5-trihydroxy-6-{[(3,4,5,6-tetrahydroxytetrahydro-2*H*-pyran-2-yl)oxy]methyl}tetrahydro-2*H*-pyran-2-yl imidoformate



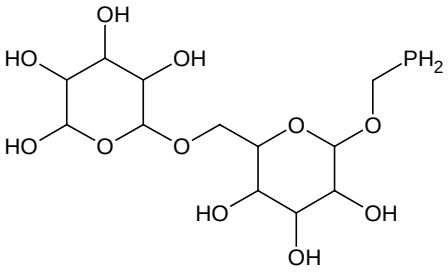
79- 6-{[6-(aminomethoxy)-3,4,5-trihydroxytetrahydro-2*H*-pyran-2-yl]methoxy}tetrahydro-2*H*-pyran-2,3,4,5-tetrol



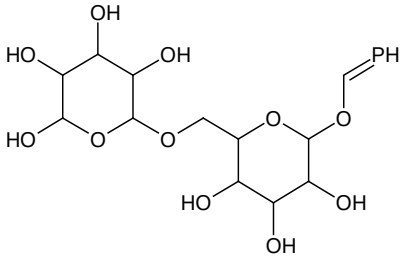
80- 6-{[6-(chloromethoxy)-3,4,5-trihydroxytetrahydro-2*H*-pyran-2-yl]methoxy}tetrahydro-2*H*-pyran-2,3,4,5-tetrol



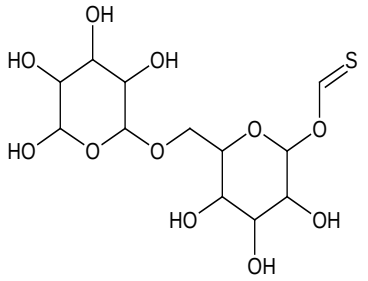
81- 6-{[6-(fluoromethoxy)-3,4,5-trihydroxytetrahydro-2*H*-pyran-2-yl]methoxy}tetrahydro-2*H*-pyran-2,3,4,5-tetrol



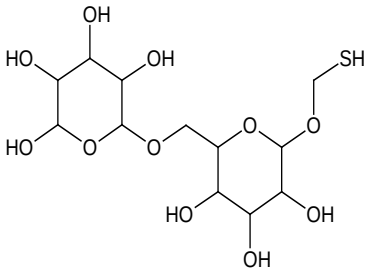
82- 6-{[3,4,5-trihydroxy-6-(phosphinomethoxy)tetrahydro-2*H*-pyran-2-yl]methoxy}tetrahydro-2*H*-pyran-2,3,4,5-tetrol



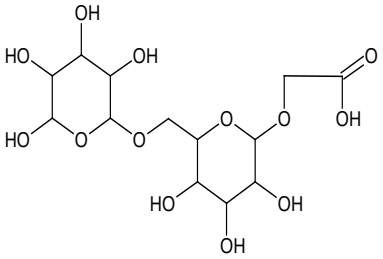
83- 6-{[3,4,5-trihydroxy-6-(phosphinidenemethoxy)tetrahydro-2*H*-pyran-2-yl]methoxy}tetrahydro-2*H*-pyran-2,3,4,5-tetrol



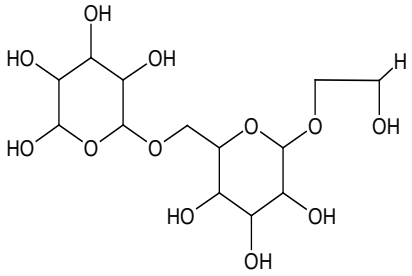
84- *O*-(3,4,5-trihydroxy-6-{[(3,4,5,6-tetrahydroxytetrahydro-2*H*-pyran-2-yl)oxy]methyl}tetrahydro-2*H*-pyran-2-yl) thioformate



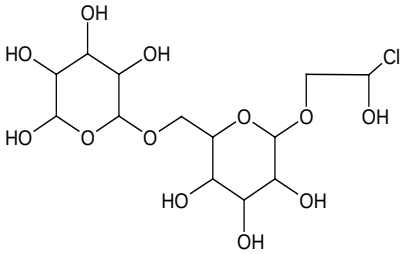
85- 6-{[3,4,5-trihydroxy-6-(mercaptomethoxy)tetrahydro-2*H*-pyran-2-yl]methoxy}tetrahydro-2*H*-pyran-2,3,4,5-tetrol



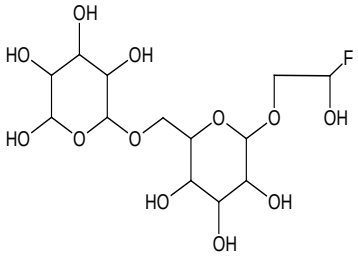
86- [(3,4,5-trihydroxy-6-{[(3,4,5,6-tetrahydroxytetrahydro-2*H*-pyran-2-yl)oxy]methyl}tetrahydro-2*H*-pyran-2-yl)oxy]acetic acid



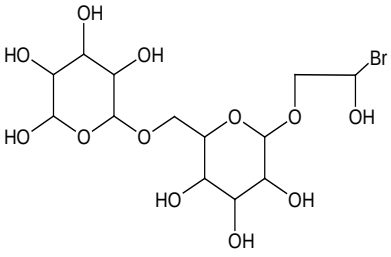
87- 6-{[3,4,5-trihydroxy-6-(2-hydroxyethoxy)tetrahydro-2*H*-pyran-2-yl]methoxy}tetrahydro-2*H*-pyran-2,3,4,5-tetrol



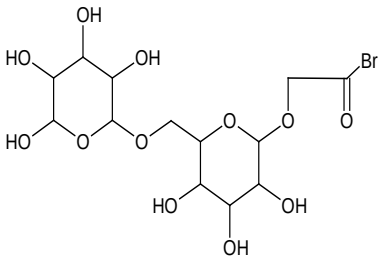
88- 6-{[6-(2-chloro-2-hydroxyethoxy)-3,4,5-trihydroxytetrahydro-2*H*-pyran-2-yl]methoxy}tetrahydro-2*H*-pyran-2,3,4,5-tetrol



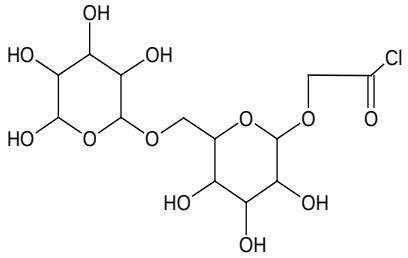
89- 6-{[6-(2-fluoro-2-hydroxyethoxy)-3,4,5-trihydroxytetrahydro-2*H*-pyran-2-yl]methoxy}tetrahydro-2*H*-pyran-2,3,4,5-tetrol



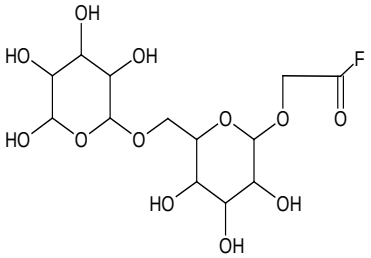
90- 6-{[6-(2-bromo-2-hydroxyethoxy)-3,4,5-trihydroxytetrahydro-2*H*-pyran-2-yl]methoxy}tetrahydro-2*H*-pyran-2,3,4,5-tetrol



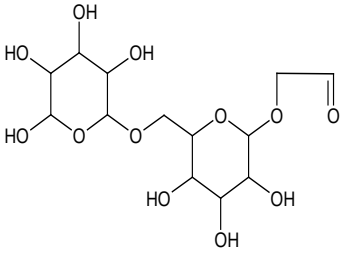
91- [(3,4,5-trihydroxy-6-{[(3,4,5,6-tetrahydroxytetrahydro-2*H*-pyran-2-yl)oxy]methyl}tetrahydro-2*H*-pyran-2-yl)oxy]acetyl bromide



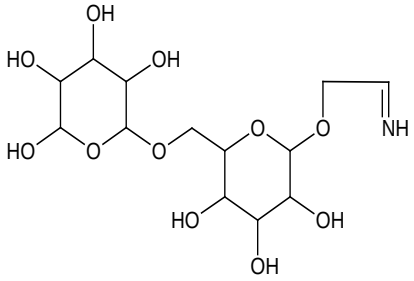
92- [(3,4,5-trihydroxy-6-{[(3,4,5,6-tetrahydroxytetrahydro-2*H*-pyran-2-yl)oxy]methyl}tetrahydro-2*H*-pyran-2-yl)oxy]acetyl chloride



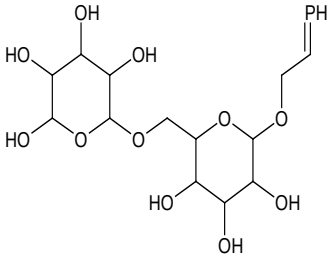
93- [(3,4,5-trihydroxy-6-{[(3,4,5,6-tetrahydroxytetrahydro-2*H*-pyran-2-yl)oxy]methyl}tetrahydro-2*H*-pyran-2-yl)oxy]acetyl fluoride



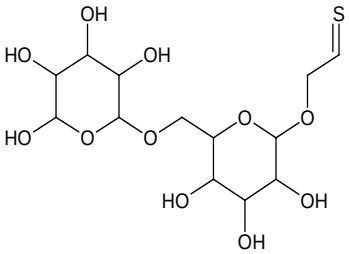
94- [(3,4,5-trihydroxy-6-{[(3,4,5,6-tetrahydroxytetrahydro-2*H*-pyran-2-yl)oxy]methyl}tetrahydro-2*H*-pyran-2-yl)oxy]acetaldehyde



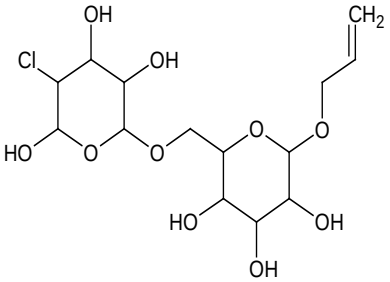
95- 6-{[3,4,5-trihydroxy-6-(2-iminoethoxy)tetrahydro-2*H*-pyran-2-yl]methoxy}tetrahydro-2*H*-pyran-2,3,4,5-tetrol



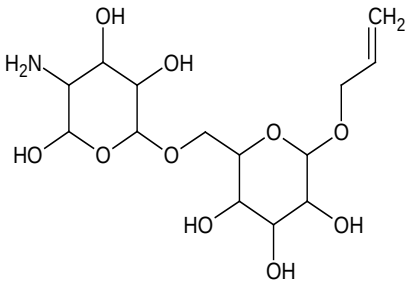
96- 6-{[3,4,5-trihydroxy-6-(2-phosphinideneethoxy)tetrahydro-2*H*-pyran-2-yl]methoxy}tetrahydro-2*H*-pyran-2,3,4,5-tetrol



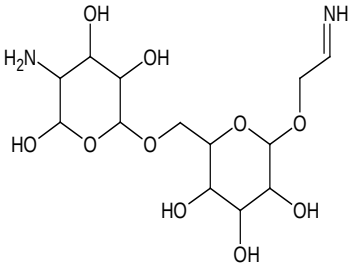
97- [(3,4,5-trihydroxy-6-{[(3,4,5,6-tetrahydroxytetrahydro-2*H*-pyran-2-yl)oxy]methyl}tetrahydro-2*H*-pyran-2-yl)oxy]ethaneth ial



98- 6-{[6-(allyloxy)-3,4,5-trihydroxytetrahydro-2*H*-pyran-2-yl]methoxy}-3-chlorotetrahydro-2*H*-pyran-2,4,5-triol



99- 6-{[6-(allyloxy)-3,4,5-trihydroxytetrahydro-2*H*-pyran-2-yl]methoxy}-3-aminotetrahydro-2*H*-pyran-2,4,5-triol



100- 3-amino-6-{[3,4,5-trihydroxy-6-(2-iminoethoxy)tetrahydro-2*H*-pyran-2-yl]methoxy}tetrahydro-2*H*-pyran-2,4,5-triol

4.5 Docking

Docking of the inhibitors with BCL2L10 was performed using GOLD 3.0.1, which is based on genetic algorithm. This program generates an ensemble of different rigid body orientations (poses) for each compound conformer within the binding pocket and then passes each molecule against a negative image of the binding site. Poses clashing with this 'bump map' are eliminated. Poses surviving the bump test are then scored and ranked with a Gaussian shape function. We defined the binding pocket using the ligand-free protein structure and a box enclosing the binding site. This box was defined by extending the size of a cocrystalized ligand by 4 Å. This dimension was considered here appropriate to allow, for instance, compounds larger than the cocrystallized ones to fit into the binding site. One unique pose for each of the best-scored compounds was saved for the subsequent steps. The compounds used for docking was converted in 3D with SILVER. To this set, the substrate corresponding to the modeled protein were added. Docking of inhibitors with the active site of protein Bcl2L10

Table 2: Individual for each ligand docked with protein in GOLD

Fitness	S(hb_ext)	S(vdw_ext)	S(hb_int)	S(int)	File name	Ligand name
35.52	8.00	22.72	0.00	-3.72	soln_1_m1_1.mol2	12301111542D'
38.50	12.67	20.18	0.00	-1.91	soln_10_m1_5.mol2	12301112062D'
40.11	13.73	21.46	0.00	-3.14	soln_11_m1_4.mol2	12301113292D'
34.37	6.00	21.79	0.00	-1.60	soln_12_m1_2.mol2	12301113302D'
32.49	6.00	20.20	0.00	-1.28	soln_13_m1_1.mol2	12301113322D'
35.12	6.04	23.07	0.00	-2.65	soln_14_m1_1.mol2	12301113332D'
34.16	8.00	20.11	0.00	-1.49	soln_15_m1_4.mol2	12301113342D'
31.05	6.00	21.43	0.00	-4.42	soln_16_m1_10.mol2	12301113362D'
33.69	6.34	23.56	0.00	-5.04	soln_17_m1_3.mol2	12301113362D'
31.69	5.99	22.52	0.00	-5.27	soln_18_m1_7.mol2'	12301113372D'
32.32	7.65	19.20	0.00	-1.72	soln_19_m1_6.mol2	12301113382D'
34.46	15.77	17.40	0.00	-5.24	soln_2_m1_5.mol2	12301111562D'
33.84	4.73	23.67	0.00	-3.44	soln_20_m1_2.mol2	12301113392D'
35.31	8.86	22.26	0.00	-4.16	soln_21_m1_9.mol2	12301113392D'
36.31	1.80	26.38	0.00	-1.76	soln_22_m1_1.mol2	12301113402D'
36.31	4.56	25.21	0.00	-2.91	soln_23_m1_9.mol2	12301113412D
35.34	4.75	24.03	0.00	-2.44	soln_24_m1_3.mol2	12301113422D
37.17	5.31	25.58	0.00	-3.31	soln_25_m1_1.mol2	12301113432D'
37.38	14.44	19.88	0.00	-4.39	soln_26_m1_1.mol2	12301113442D'
33.56	13.20	19.41	0.00	-6.34	soln_3_m1_5.mol2	12301112002D'
33.63	12.78	19.41	0.00	-5.84	soln_4_m1_7.mol2'	12301112012D'
29.81	11.39	18.79	0.00	-7.41	soln_5_m1_4.mol2	12301112022D'
33.94	14.15	20.65	0.00	-8.60	soln_6_m1_6.mol2	12301112032D'
32.63	13.71	21.48	0.00	-10.61	soln_7_m1_1.mol2	12301112042D
33.42	7.53	20.01	0.00	-1.63	soln_8_m1_6.mol2	12301112042D'
33.60	1.25	25.64	0.00	-2.91	soln_9_m1_2.mol2	12301112052D'
33.42	6.60	25.53	0.00	-8.29	soln_100_m1_5.mol2	12301118262D'
32.96	12.26	18.23	0.00	-4.36	soln_27_m1_3.mol2	12301113442D'
31.79	11.26	19.04	0.00	-5.65	soln_28_m1_6.mol2	12301113452D'
34.95	20.42	17.21	0.00	-9.14	soln_29_m1_8.mol2	12301113462D'
37.88	16.83	20.10	0.00	-6.58	soln_30_m1_5.mol2	12301113472D'

36.66	17.12	18.08	0.00	-5.33	soln_31_m1_7.mol2	12301113492D'
34.33	16.00	19.56	0.00	-8.57	soln_32_m1_5.mol2	12301113502D'
31.93	12.84	18.76	0.00	-6.71	soln_33_m1_4.mol2	12301113512D'
32.11	6.80	21.12	0.00	-3.73	soln_34_m1_3.mol2	12301113512D
33.01	13.37	17.09	0.00	-3.85	soln_35_m1_3.mol2	12301113522D'
29.48	4.95	19.92	0.00	-2.87	soln_36_m1_2.mol2	12301113532D'
34.92	13.44	18.53	0.00	-4.00	soln_37_m1_2.mol2	12301113542D'
31.54	5.61	21.58	0.00	-3.75	soln_38_m1_4.mol2	12301113552D'
36.90	13.56	21.11	0.00	-5.69	soln_39_m1_2.mol2	12301113562D'
35.09	20.28	15.10	0.00	-5.96	soln_40_m1_8.mol2	12301113572D'
33.48	13.28	20.36	0.00	-7.80	soln_41_m1_10.mol2	12301113582D'
36.39	12.33	20.71	0.00	-4.42	soln_42_m1_8.mol2	12301113592D'
31.54	0.00	25.86	0.00	-4.02	soln_43_m1_1.mol2	12301114002D'
29.87	5.25	19.92	0.00	-2.77	soln_44_m1_1.mol2	12301114012D'
34.82	6.45	22.34	0.00	-2.35	soln_45_m1_8.mol2	12301114022D'
31.96	5.85	23.76	0.00	-6.56	soln_46_m1_4.mol2	12301114032D'
36.65	12.35	18.64	0.00	-1.34	soln_47_m1_5.mol2	12301114042D'
35.95	12.17	18.23	0.00	-1.27	soln_48_m1_2.mol2	12301114052D'
36.62	16.13	18.14	0.00	-4.46	soln_49_m1_7.mol2	12301114062D'
37.22	11.44	19.64	0.00	-1.22	soln_50_m1_4.mol2	12301114082D'
35.06	10.47	19.20	0.00	-1.81	soln_51_m1_7.mol2	12301114092D'
32.03	9.03	19.13	0.00	-3.30	soln_52_m1_1.mol2	12301114102D'
35.17	10.94	19.77	0.00	-2.95	soln_53_m1_8.mol2	12301114122D'
36.82	13.79	18.18	0.00	-1.97	soln_54_m1_10.mol2	12301114142D'
38.77	14.23	19.68	0.00	-2.51	soln_55_m1_8.mol2	12301114142D'
37.42	14.14	18.27	0.00	-1.84	soln_56_m1_2.mol2	12301114152D'
39.50	14.20	20.91	0.00	-3.44	soln_57_m1_2.mol2	12301114162D'
36.81	12.54	20.33	0.00	-3.68	soln_58_m1_4.mol2	12301114172D'
36.12	6.21	22.72	0.00	-1.32	soln_59_m1_1.mol2	12301114182D'
32.86	11.99	16.72	0.00	-2.11	soln_60_m1_3.mol2	12301114192D'
32.78	6.68	19.94	0.00	-1.32	soln_61_m1_1.mol2	12301114202D'
33.21	7.48	20.14	0.00	-1.96	soln_62_m1_7.mol2	12301114232D'
40.20	4.77	28.25	0.00	-3.40	soln_63_m1_2.mol2	12301114242D'
40.54	1.41	31.62	0.00	-4.34	soln_64_m1_2.mol2	12301114262D'
29.06	0.00	23.87	0.00	-3.76	soln_65_m1_3.mol2	12301114282D'
27.09	4.71	19.84	0.00	-4.90	soln_66_m1_8.mol2	12301114282D'

31.23	6.00	20.67	0.00	-3.19	soln_67_m1_2.mol2	12301114292D'
28.44	5.83	19.76	0.00	-4.56	soln_68_m1_9.mol2	12301114302D'
29.19	4.60	22.33	0.00	-6.11	soln_69_m1_2.mol2	12301114312D'
30.03	6.00	22.23	0.00	-6.53	soln_70_m1_3.mol2	12301114332D'
33.61	7.98	24.04	0.00	-7.43	soln_71_m1_4.mol2	12301114332D'
36.59	16.90	21.59	0.00	-9.99	soln_72_m1_1.mol2	12301114352D'
37.83	6.85	27.75	0.00	-7.18	soln_73_m1_2.mol2	12301114362D'
34.69	17.99	20.88	0.00	-12.00	soln_74_m1_9.mol2	12301114562D'
39.49	12.87	24.10	0.00	-6.52	soln_75_m1_3.mol2	12301114572D'
33.27	9.66	21.59	0.00	-6.08	soln_76_m1_2.mol2	12301114582D'
36.45	9.01	22.84	0.00	-3.97	soln_77_m1_4.mol2	12301115002D'
40.16	11.58	23.22	0.00	-3.35	soln_78_m1_5.mol2	12301115002D'
29.54	15.95	15.91	0.00	-8.29	soln_79_m1_1.mol2	12301115012D'
30.94	21.25	17.00	0.00	-13.68	soln_80_m1_10.mol2	12301115022D'
28.05	11.46	18.28	0.00	-8.55	soln_81_m1_9.mol2	12301115032D'
31.12	10.35	23.50	0.00	-11.54	soln_82_m1_7.mol2	12301115032D'
34.88	19.91	15.40	0.00	-6.21	soln_83_m1_4.mol2	12301115042D'
35.60	4.18	28.11	0.00	-7.24	soln_84_m1_3.mol2	12301115052D'
35.73	12.36	23.53	0.00	-8.98	soln_85_m1_4.mol2	12301115062D'
30.57	12.08	21.85	0.00	-11.56	soln_86_m1_6.mol2	12301115082D'
32.24	12.26	21.36	0.00	-9.39	soln_87_m1_10.mol2	12301118152D
31.71	9.99	25.01	0.00	-12.66	soln_88_m1_5.mol2	12301118132D'
33.34	10.58	23.50	0.00	-9.56	soln_89_m1_3.mol2	12301118162D'
31.03	9.51	24.97	0.00	-12.82	soln_90_m1_7.mol2	12301118162D'
33.14	8.63	24.67	0.00	-9.42	soln_91_m1_4.mol2	12301118182D'
31.19	15.12	19.63	0.00	-10.92	soln_92_m1_10.mol2	12301118192D'
31.81	5.85	26.04	0.00	-9.84	soln_93_m1_5.mol2	12301118202D'
38.77	9.23	25.72	0.00	-5.82	soln_94_m1_4.mol2	12301118212D'
37.04	10.69	24.02	0.00	-6.67	soln_95_m1_9.mol2	12301118222D'
39.22	6.43	28.65	0.00	-6.61	soln_96_m1_6.mol2	12301118222D'
41.17	11.28	28.86	0.00	-9.80	soln_97_m1_6.mol2	12301118232D'
35.71	7.23	27.35	0.00	-9.12	soln_98_m1_10.mol2	12301118242D'
33.51	13.94	21.02	0.00	-9.34	soln_99_m1_2.mol2	12301118252D'

4.5.1 Docking of BCL2L10 with inhibitors

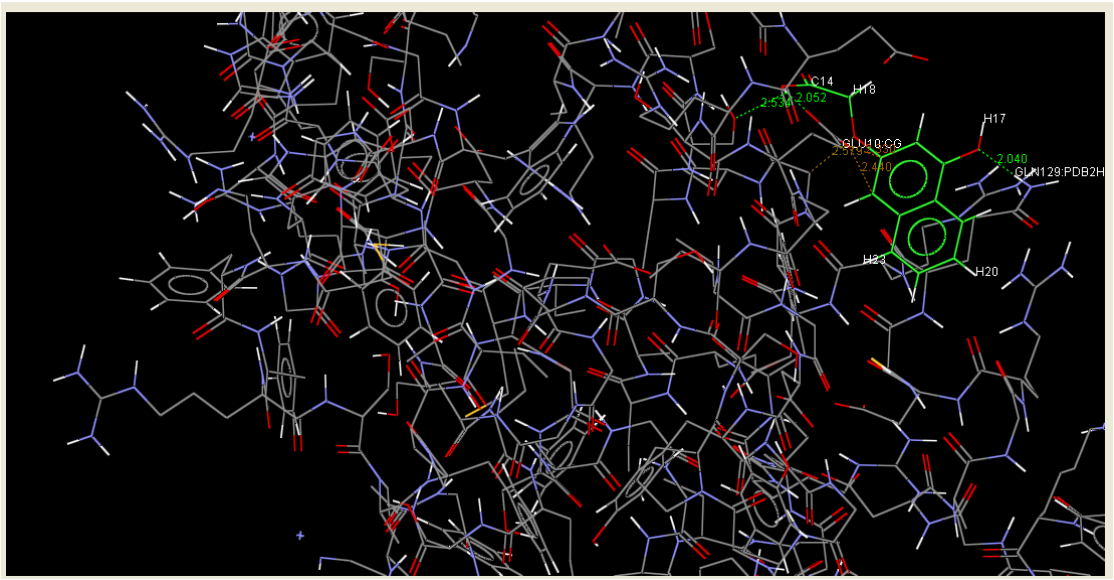


Figure 15: Docking of Bcl2L10 with drug (No 1) on level 35.52

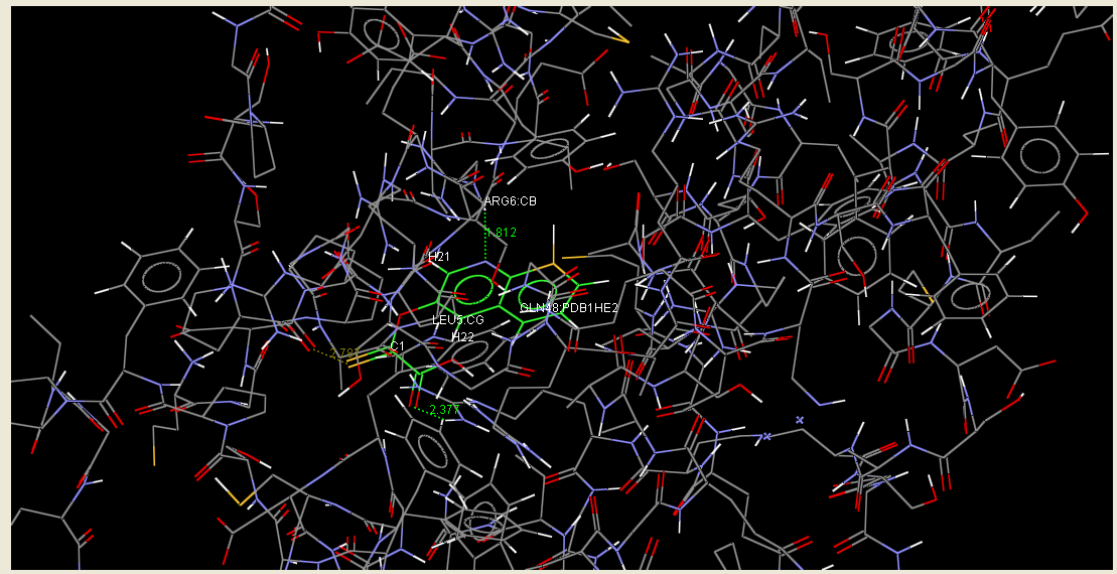


Figure 16: Docking of Bcl2L10 with drug (No 22) on level 36.31

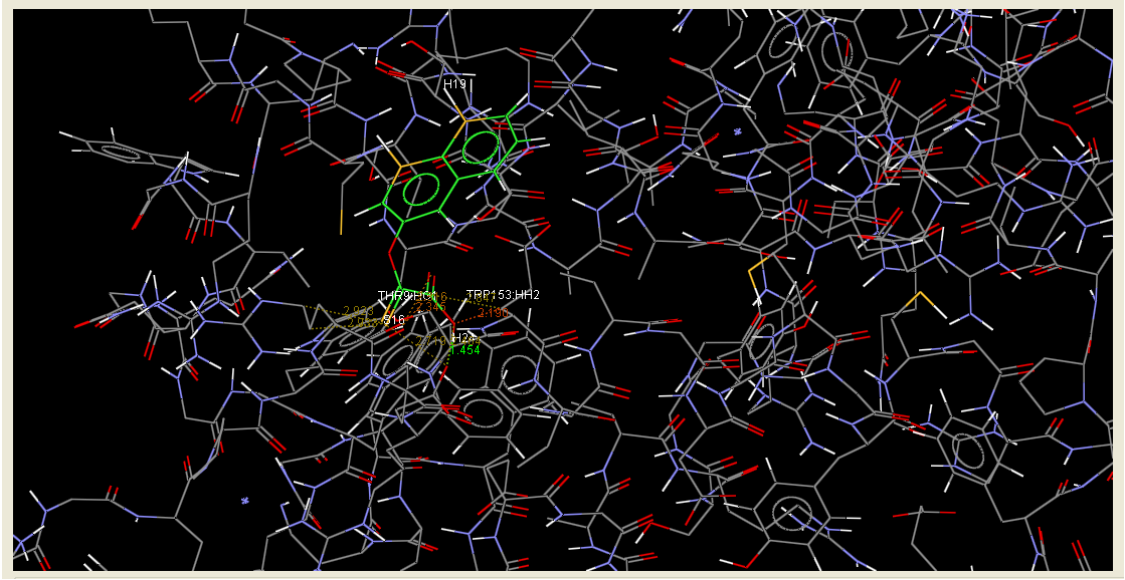


Figure 17: Docking of Bcl2L10 with drug (No 23) on level 36.31

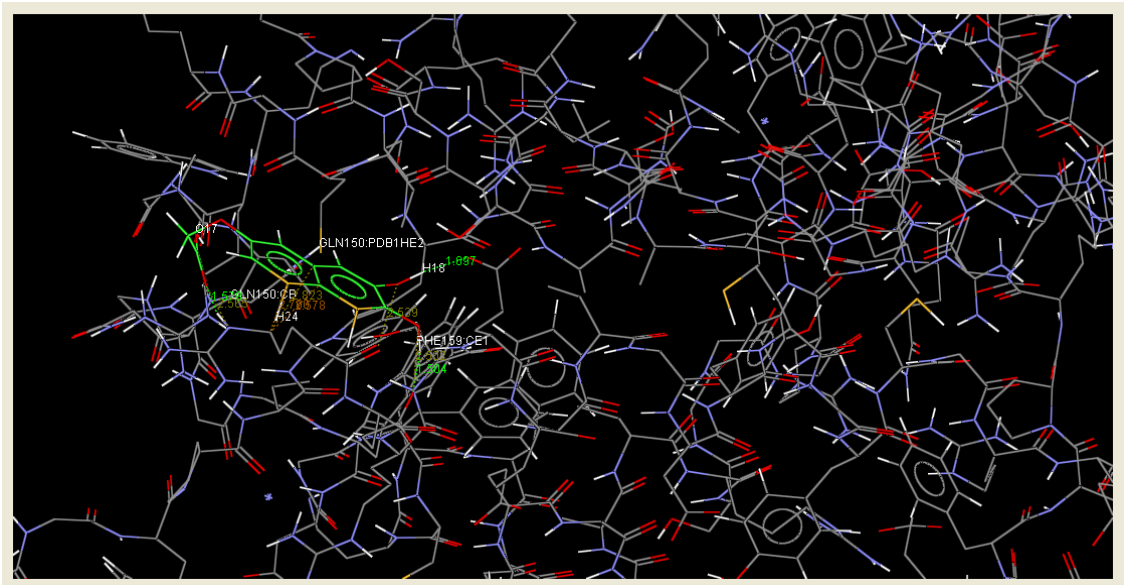


Figure 18: Docking of Bcl2L10 with drug (No 26) on level 37.38

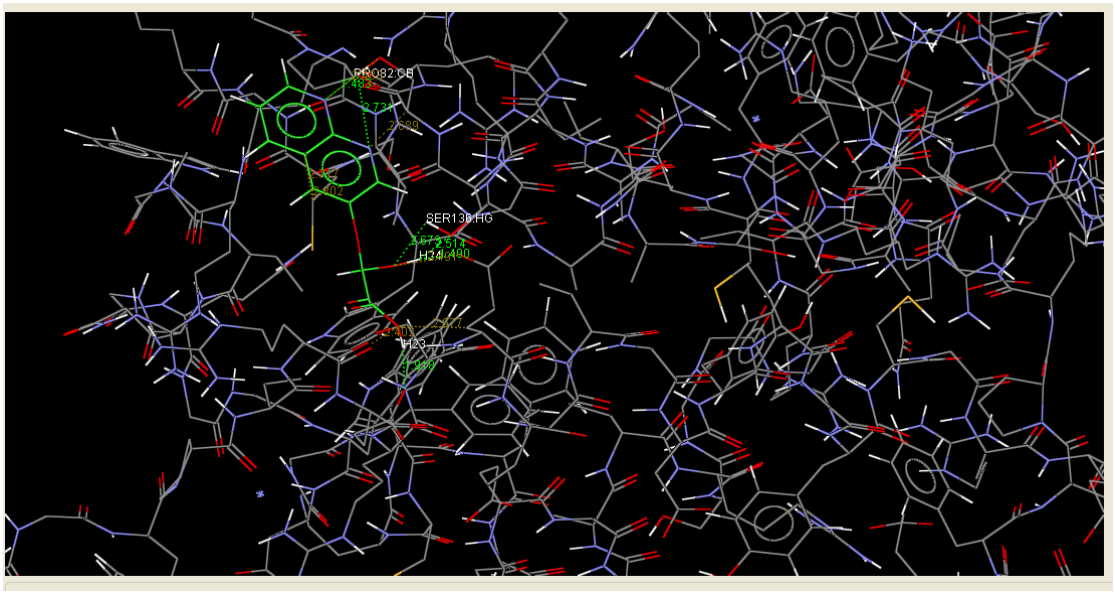


Figure 19: Docking of Bcl2L10 with drug (No 10) on level 38.50

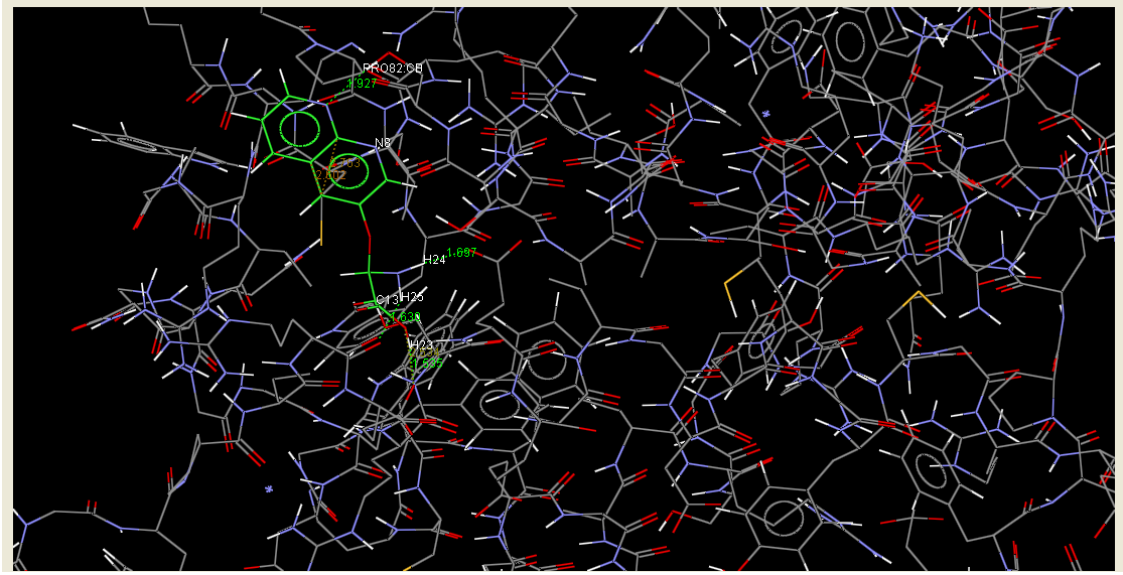


Figure 20: Docking of Bcl2L10 with drug (No 11) on level 40.11

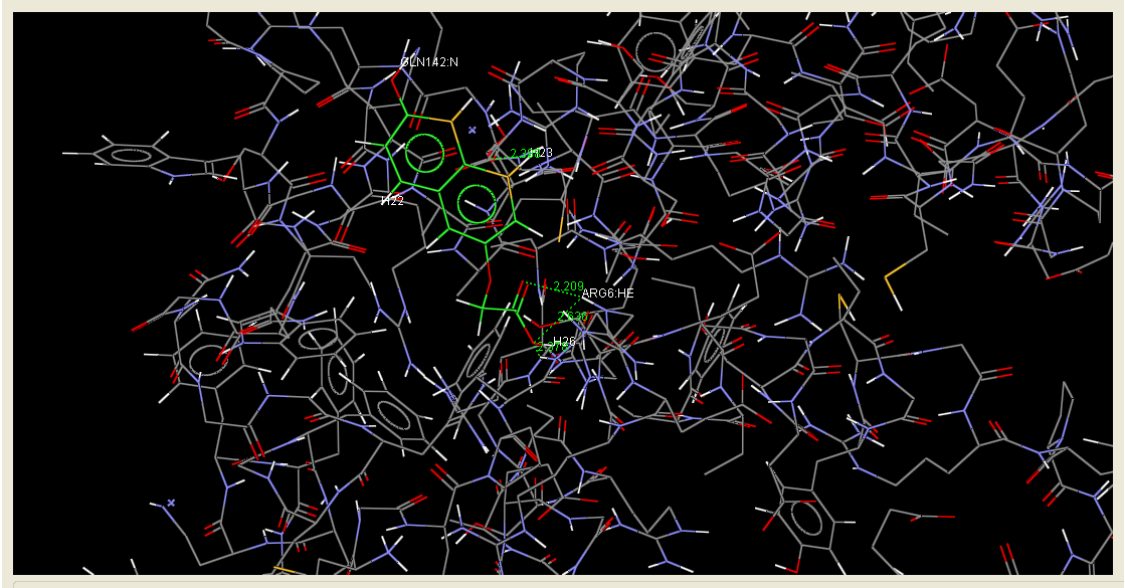


Figure 21: Docking of Bcl2L10 with drug (No 25) on level 37.17

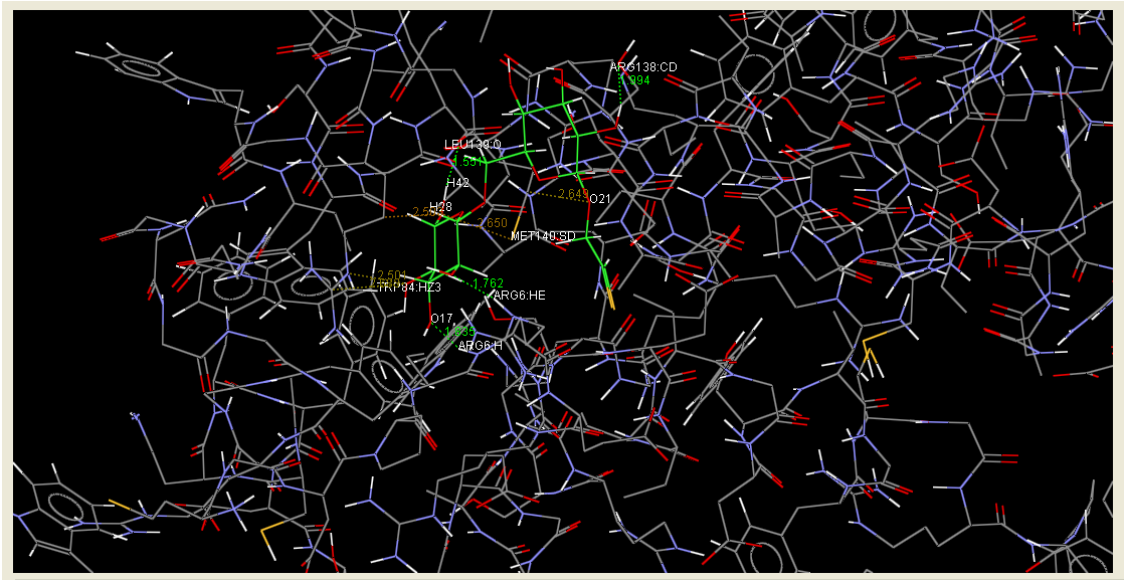


Figure 22: Docking of Bcl2L10 with drug (No 97) on level 41.17

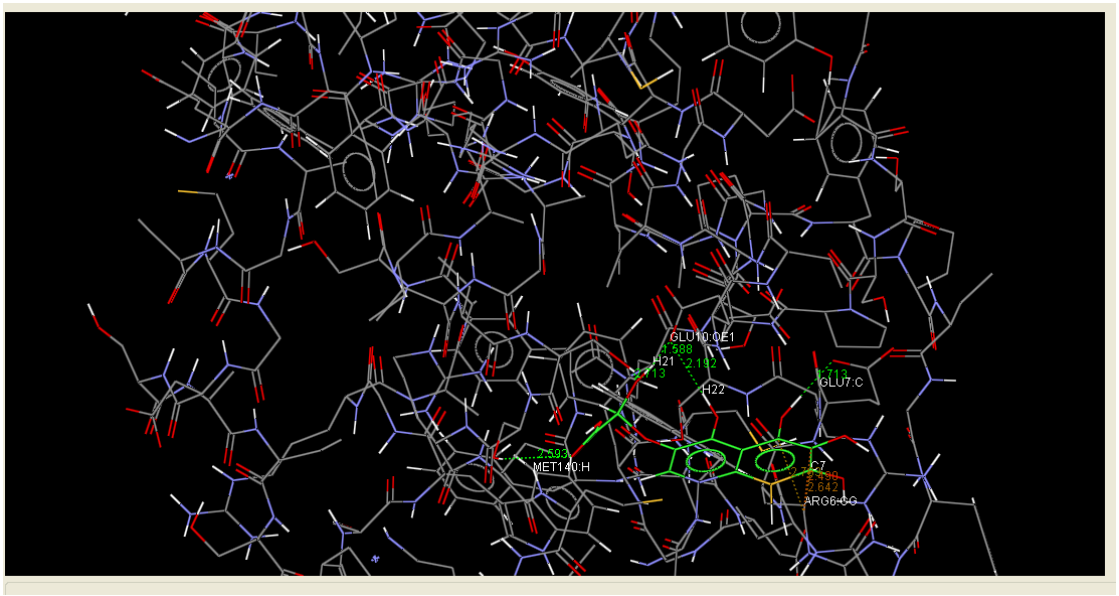


Figure 23: Docking of Bcl2L10 with drug (No 30) on level 37.88

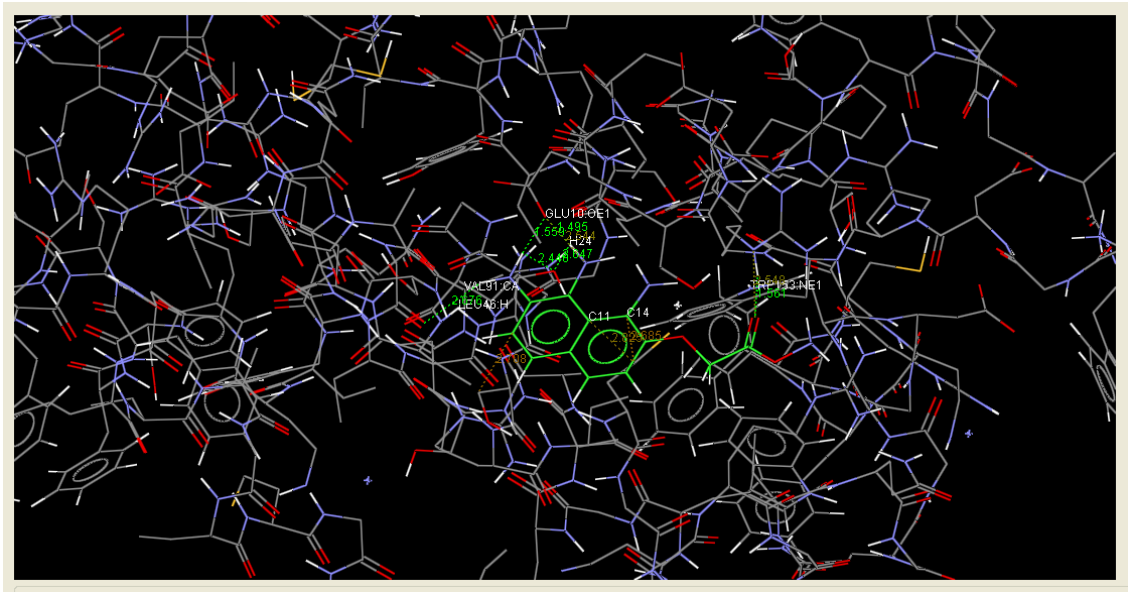


Figure 24: Docking of Bcl2L10 with drug (No 55) on level 38.77

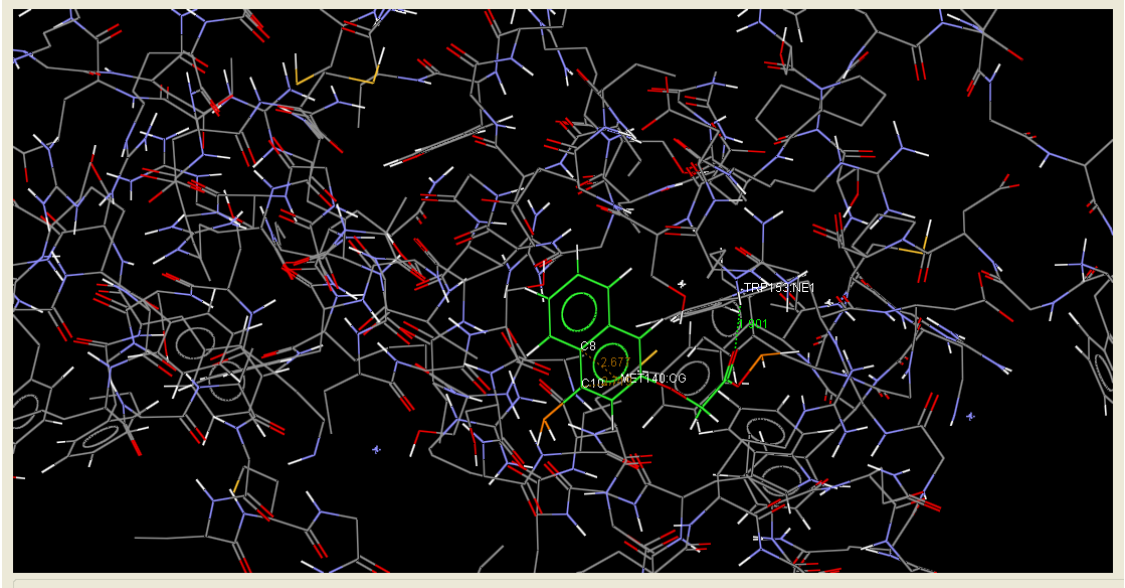


Figure 25: Docking of Bcl2L10 with drug (No 63) on level 40.20

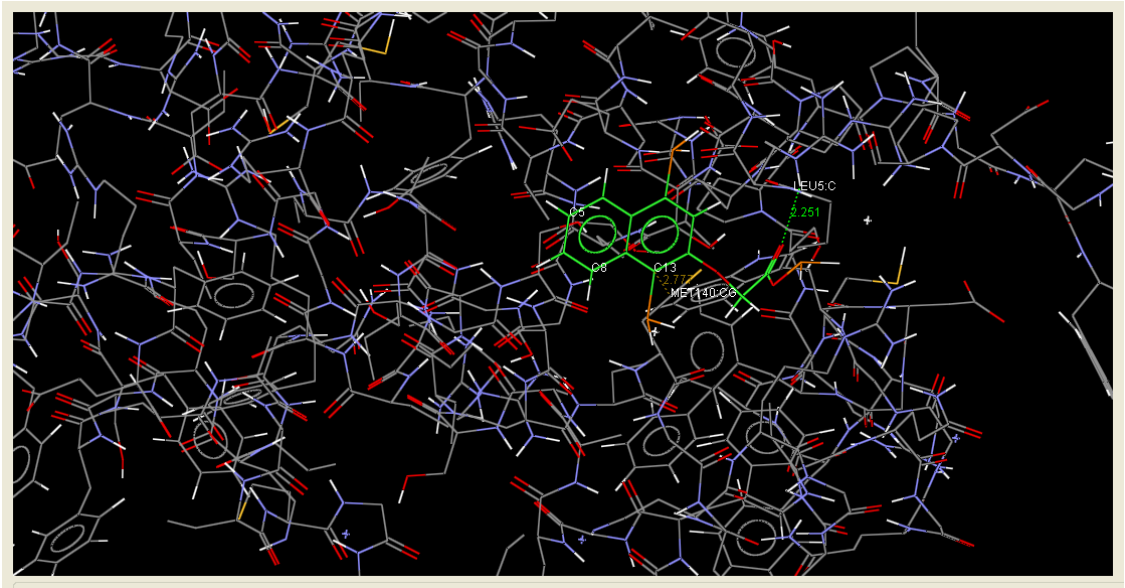


Figure 26: Docking of Bcl2L10 with drug (No 64) on level 40.54

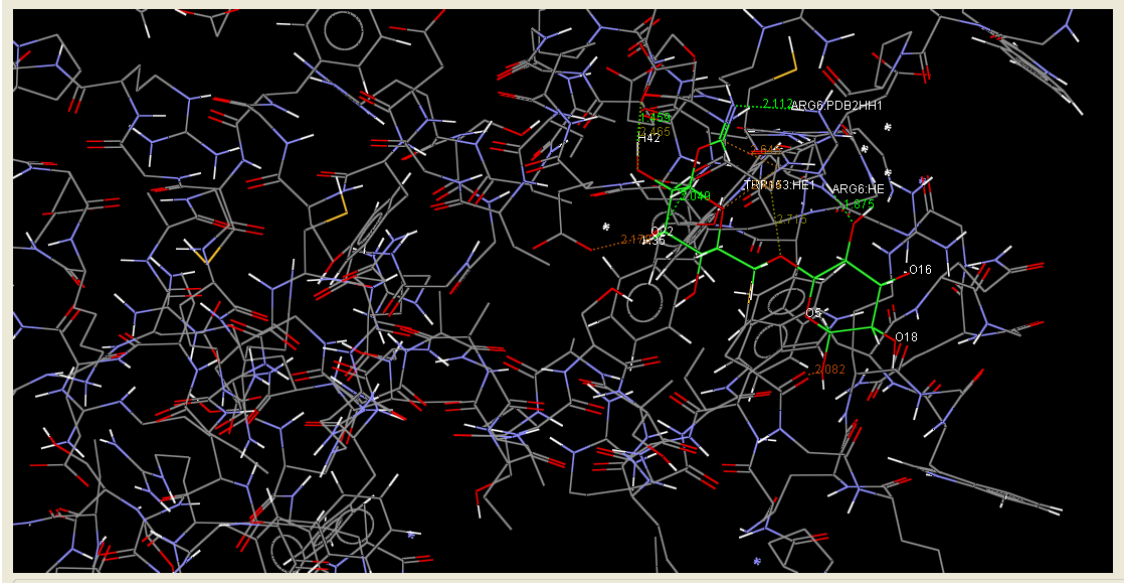


Figure 27: Docking of Bcl2L10 with drug (No 78) on level 40.16

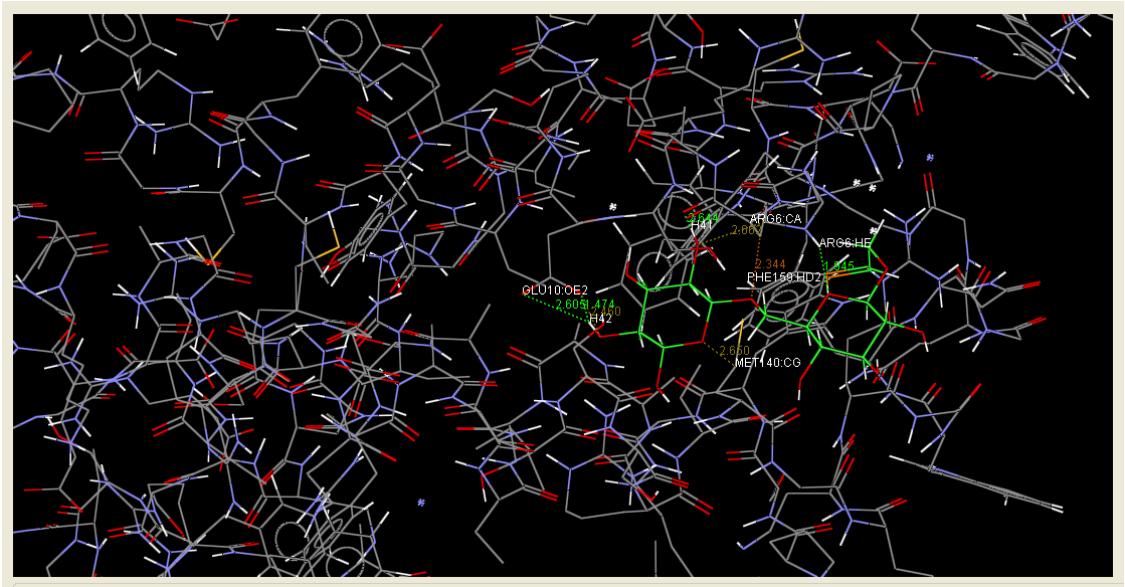


Figure 28: Docking of Bcl2L10 with drug (No 96) on level 39.22

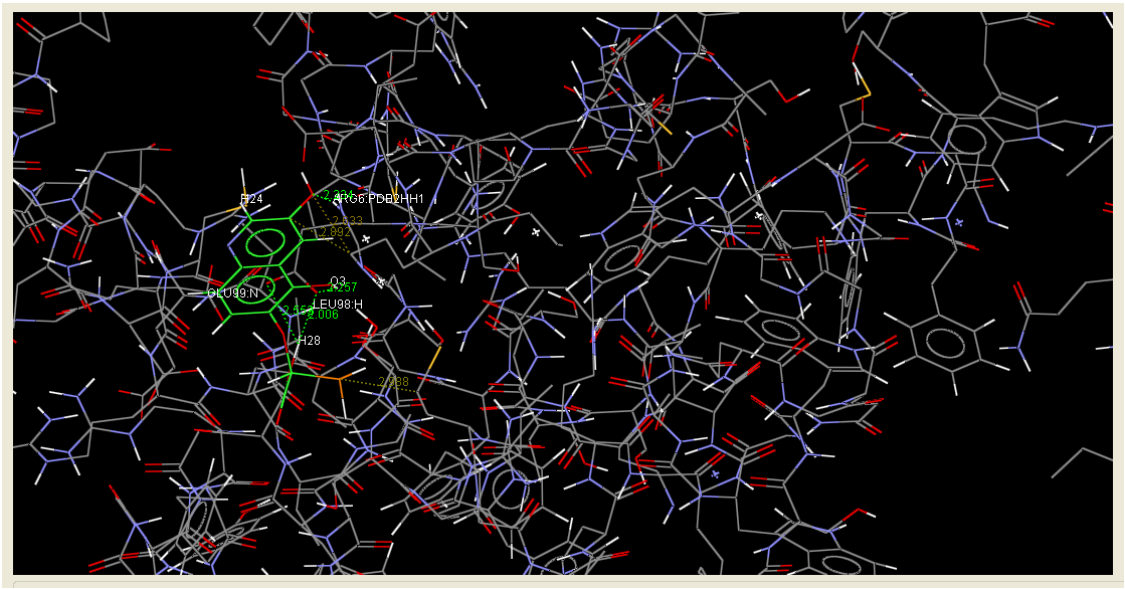


Figure 29: Docking of Bcl2L10 with drug (No 42) on level 36.39

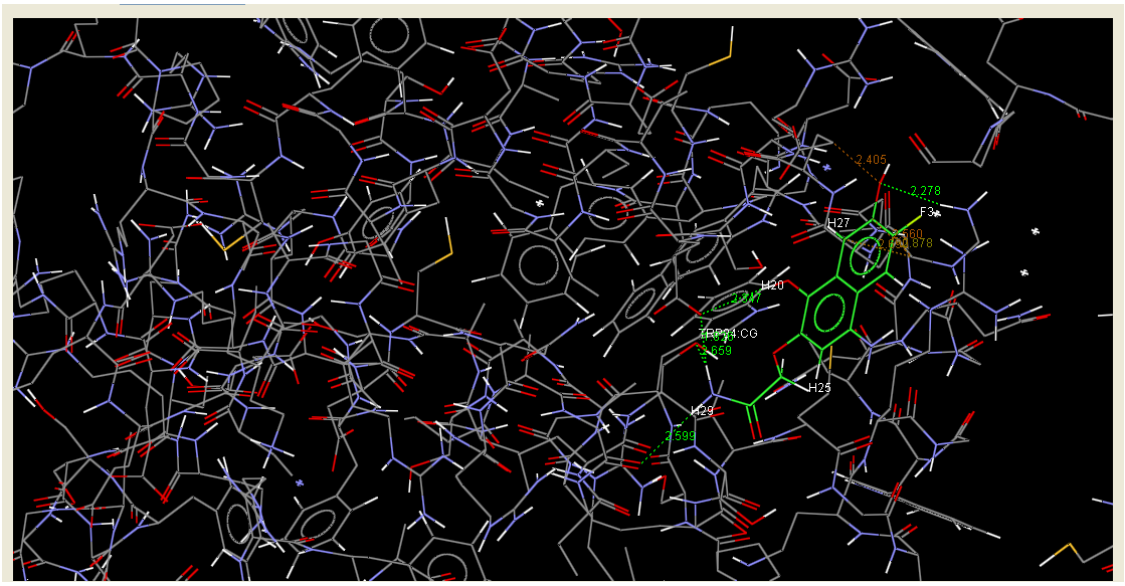


Figure 30: Docking of Bcl2L10 with drug (No 49) on level 36.62

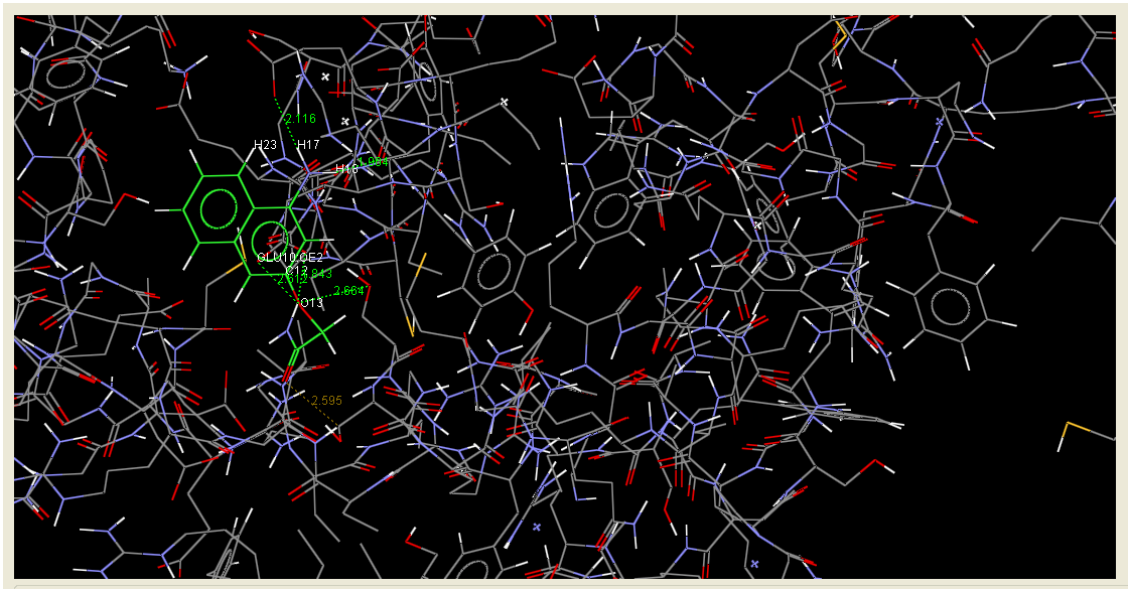


Figure 31: Docking of Bcl2L10 with drug (No 50) on level 37.22

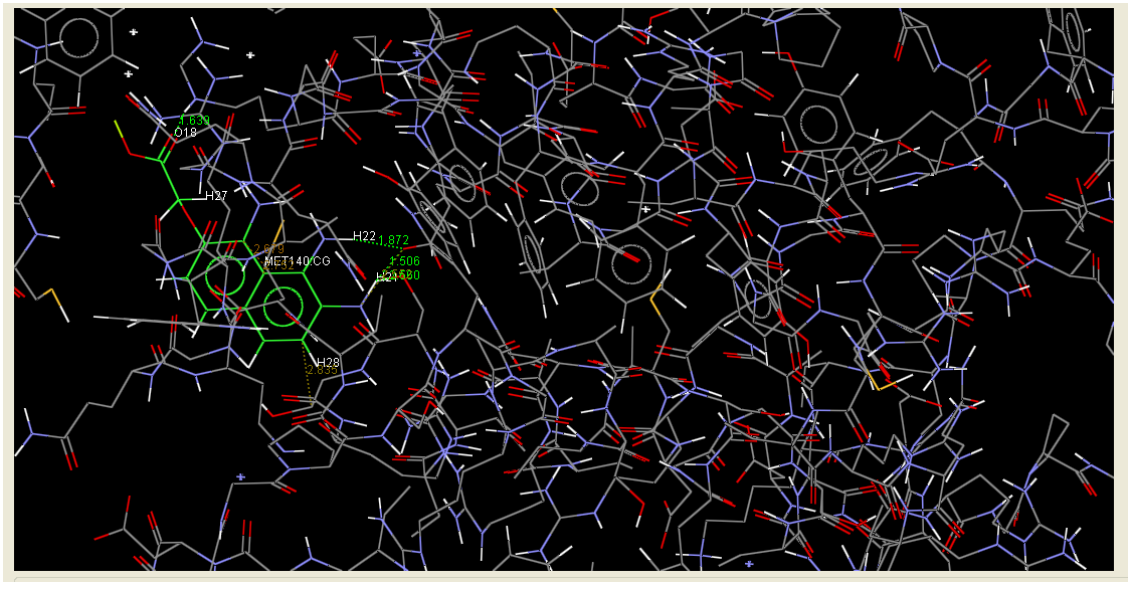


Figure 32: Docking of Bcl2L10 with drug (No 56) on level 37.42

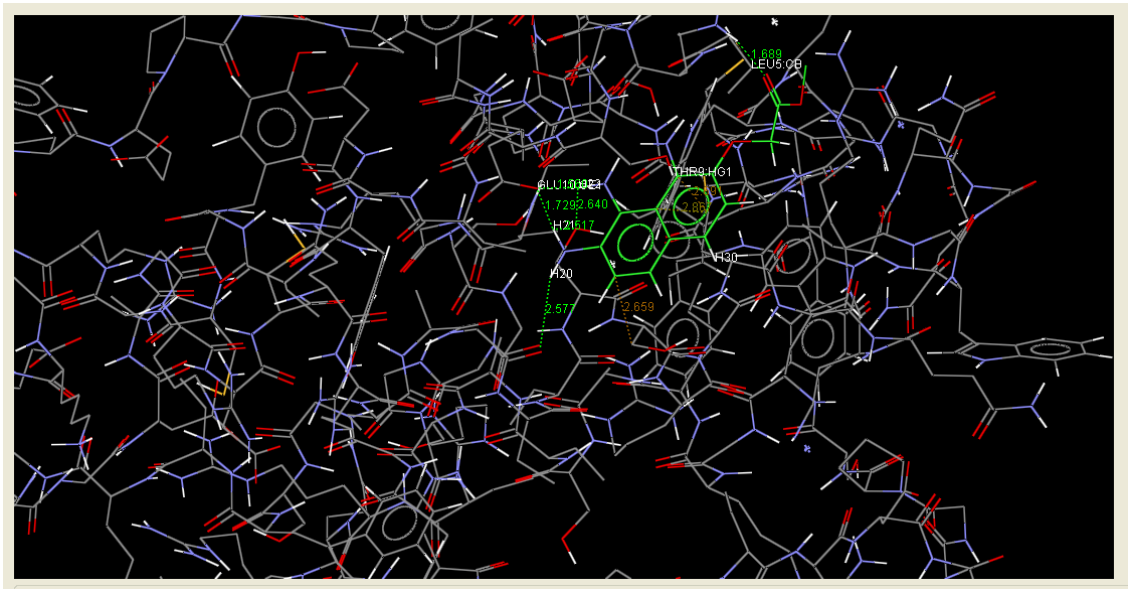


Figure 33: Docking of Bcl2L10 with drug (No 57) on level 39.50

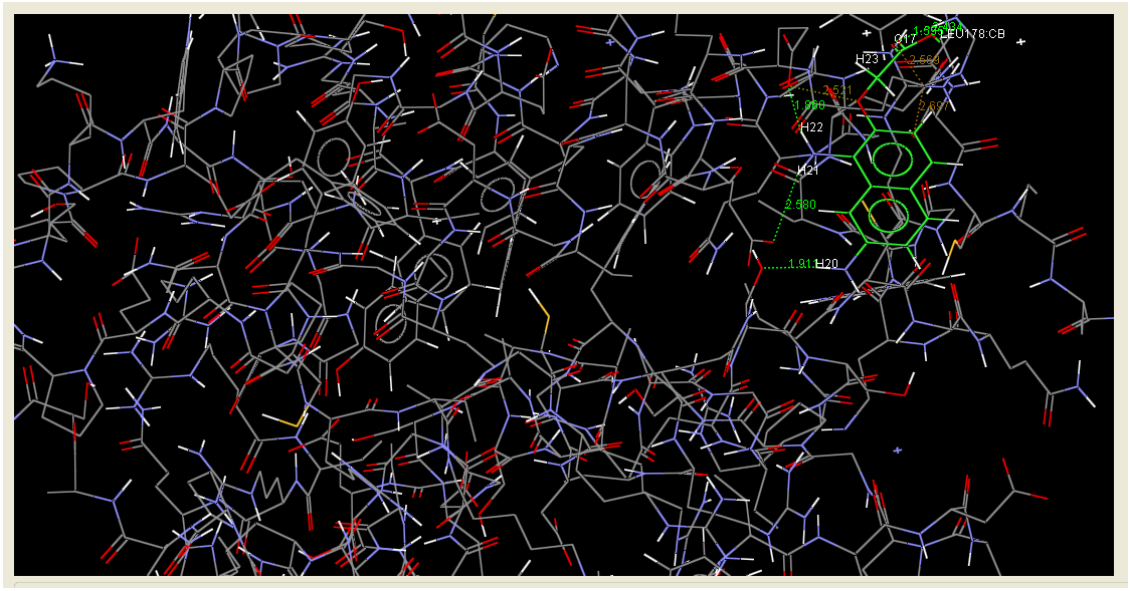


Figure 34: Docking of Bcl2L10 with drug (No 58) on level 36.81

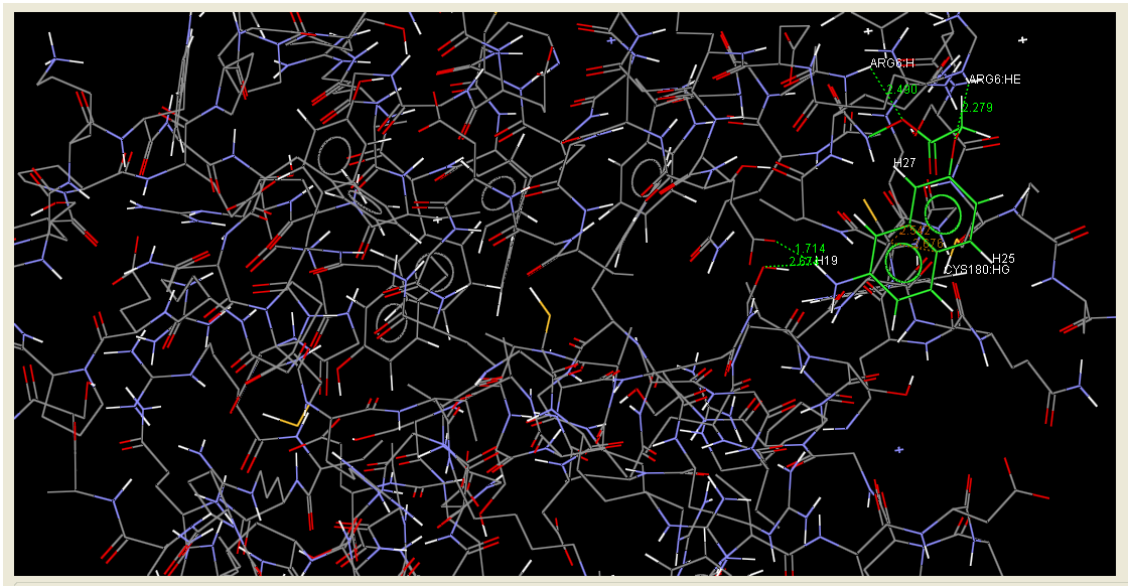


Figure 35: Docking of Bcl2L10 with drug (No 59) on level 36.12

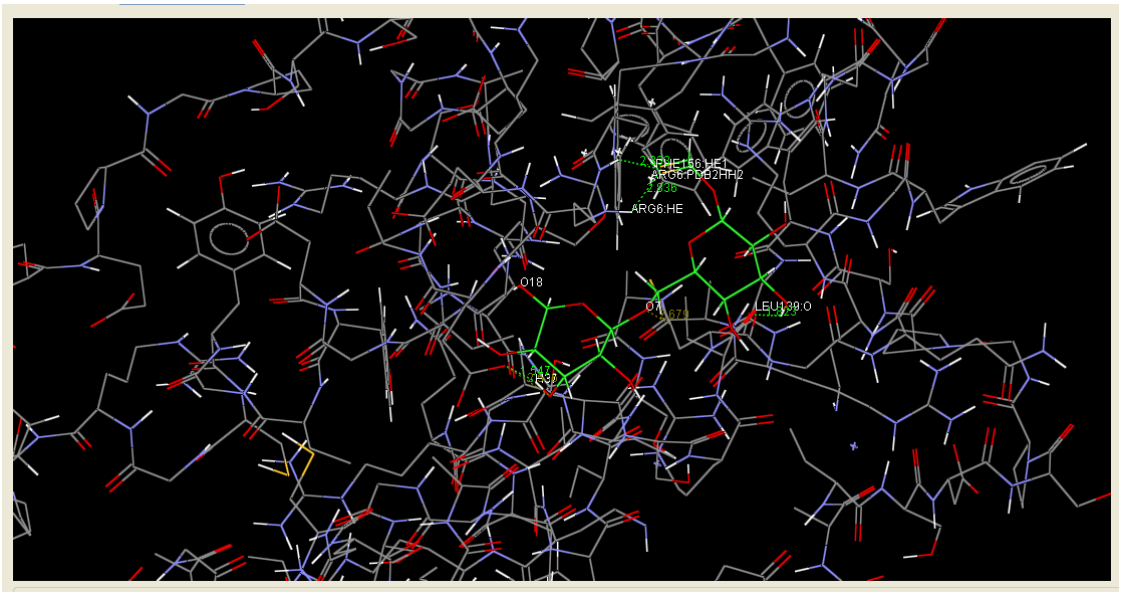


Figure 36: Docking of Bcl2L10 with drug (No 84) on level 35.60

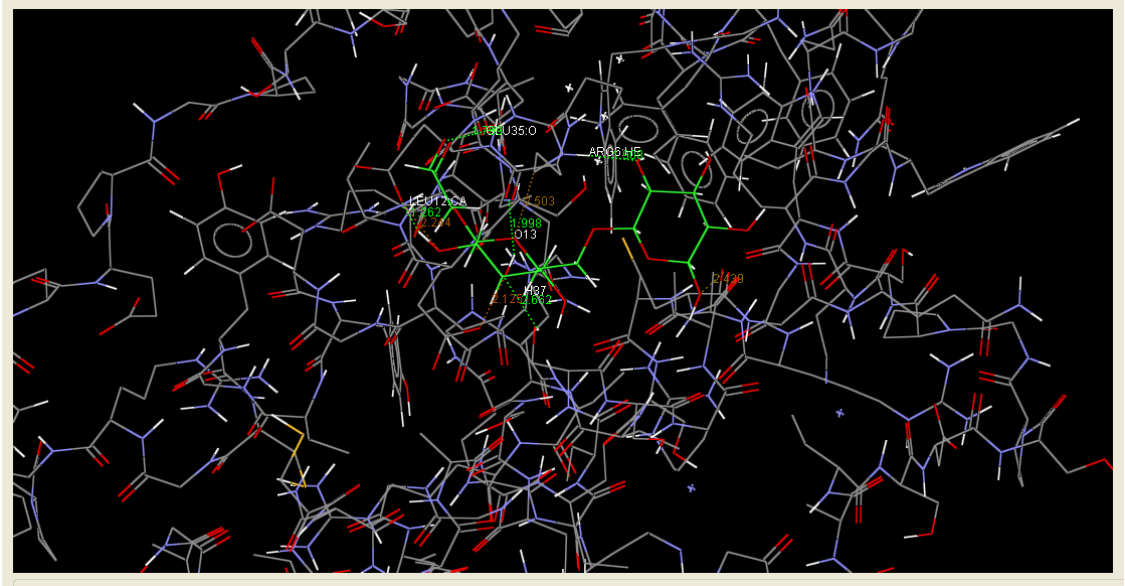


Figure 37: Docking of Bcl2L10 with drug (No 94) on level 38.77

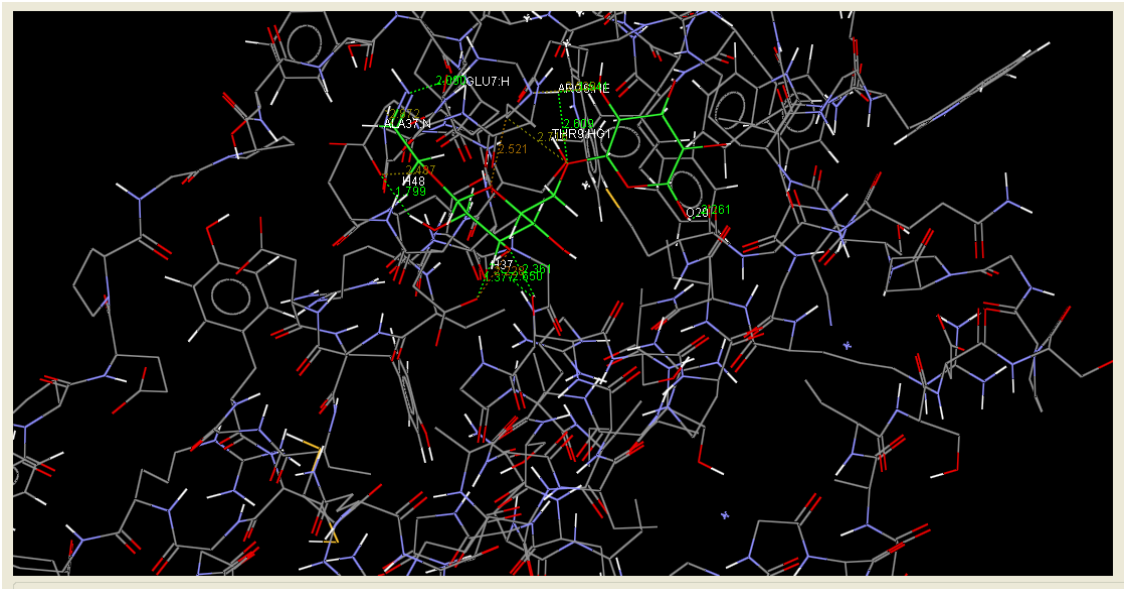


Figure 38: Docking of Bcl2L10 with drug (No 95) on level 37.04

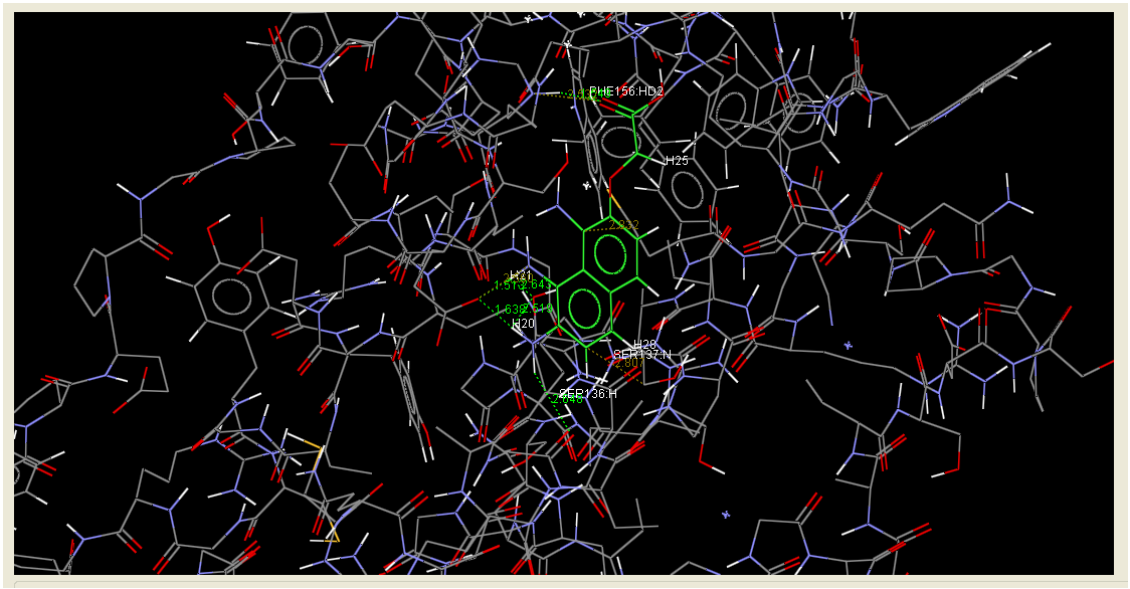


Figure 39: Docking of Bcl2L10 with drug (No 54) on level 36.82

Chapter-5 Discussion

Bioinformatics involves discovery, development and implementation of computational algorithms and software tools that facilitate an understanding of the biological processes with the goal to serve primarily healthcare sectors with several spin-offs. Cancer is a danger disease, have salient symptoms at initiatory stage and have the long latent periods. Proteins in the Bcl-2 family are central regulators of programmed cell death and members that inhibit apoptosis reductions and Bcl-2 increase sensitivity to anticancer drugs. The development of inhibitors of these proteins as potential anti-cancer therapeutics has been previously explored, but obtaining potent small-molecule inhibitors has proved difficult owing to the necessity of targeting a protein–protein interaction.

The purpose of this study is to identify the receptor having more binding energy towards the Bcl2L10. Only protein 2KUA (chain A) has a high level of coverage (85%), total score (137) and identity (48%) of the reference protein with the Bcl2L10 is 48%. It is showed from the result of multiple sequence alignment (Clustal X) between template (2KUA 165 aa) and query (Bcl2 L10 194 aa) that the alignment score was 481. Also nb atoms involved was 157 and RSM 0.39 A.̊The objective of this project is to generate the protein 3D Bcl2L10 and drugs targeted at specific docking of protein with inhibitor solution of the structure of antiapoptotic Bcl-2 family proteins has led to the design of novel small molecule inhibitors. From the 20 pdb structures generated by the MODELLER we have selected least objective function structure for better inhibitory results. This structure was submitted to Ramachandran plot and from the Table 1 it was identified that the generated model was reliable. This 3D structure was again improved through Molecular dynamics studies and finally active site residues were identified using Castp server. The residues, PHE5, LYS7, CYS9, TYR10GLY11,TYR12, CYS13, ILE14, PHE28, LEU30, TYR31, ALA61, VAL62, THR63, PHE82, VAL230, LEU281 were involved in active site of the BCL2L10.

Although many such molecules have been synthesized, rigorous verification of their specificity has often been lacking. Defects in apoptotic pathways can promote cancer cell survival and also confer resistance to anti-neoplastic drugs. One pathway being targeted for anti-neoplastic therapy is the anti-apoptotic B-cell lymphoma-2 (Bcl-2) family of proteins (Bcl-2, Bcl-X_L, Bcl-w, Mcl-1, Bfl1/A-1, and Bcl-B) that bind to and inactivate BH3-domain pro-apoptotic

proteins. According to data presented in results, it was observed that drugs inhibited Bcl2L10 in the following order

1-[(3,4,5-trihydroxy-6-{[(3,4,5,6-tetrahydroxytetrahydro-2*H*-pyran-2-yl)oxy]methyl}tetrahydro-2*H*-pyran-2-yl)oxy]ethanethial

2-{2-[2-oxo-2-(phosphinooxy)ethoxy]naphthalene-1,4-diyl}diphosphine

3-{3-[2-oxo-2-(phosphinooxy)ethoxy]-1-naphthyl}phosphine

4-3,4,5-trihydroxy-6-{[(3,4,5,6-tetrahydroxytetrahydro-2*H*-pyran-2yl)oxy]methyl} tetrahydro-2*H*-pyran-2-yl imidoformate

5- Amino(1,8-naphthyridin-3-yloxy)acetic acid

As mentioned by the data above, it suggests form testing of these drugs at experimental investigation. There is now great excitement in the field of cancer research with the development of a Bcl-2 inhibitor, ABT-737, by Abbott laboratories with the potential to overcome this block to apoptosis. [(3,4,5-trihydroxy-6-{[(3,4,5,6-tetrahydroxytetrahydro-2*H*-pyran-2-yl)oxy]methyl} tetrahydro-2*H*-pyran-2-yl)oxy] ethanethial showed better docking energy value (30.5K.cal/mol)than synthesized ABT-737 (20.8K.cal/mol). And while positive response in stimulating apoptosis has been seen in laboratory tests in human small cell.

Chapter-6 Conclusion

Despite the great interest of efforts for cancer treatment still remain the needs to investigate more because no much effect of actively arresting this disease. In this work, we have constructed a 3D model of Bcl2L10, using the ICM Molsoft software and obtained a refined model after energy minimization. The final refined model was further assessed by ERRAT and PROCHECK program, and the results show that this model is reliable. Only protein2KUA (chain A) has a high level of coverage (85%), total score (137) and identity (48%) of the reference protein with the Bcl2L10are 48%. Docking results indicate that conserved amino-acid residues Bcl2 mainly play an important role in maintaining a functional conformation and are directly involved in donor substrate binding. The interaction between the domain and the inhibitors proposed in this study are useful for understanding the potential mechanism of domain and the inhibitor binding.

As is well known, hydrogen bonds play important role for the structure and function of biological molecules. In this study it was found that,* [(3,4,5-trihydroxy-6-{[(3,4,5,6-tetrahydroxytetrahydro-2*H*-pyran-2-yl)oxy]methyl}tetrahydro-2*H*-pyran-2-yl)oxy]ethanethial, *{2-[2-oxo-2-(phosphinooxy)ethoxy] naphthalene-1,4-diyl}diphosphine, *{3-[2-oxo-2-(phosphinooxy)ethoxy]-1-naphthyl}phosphine, * 3,4,5-trihydroxy-6-{[(3,4,5,6-tetra hydroxytetrahydro-2*H*-pyran2yl)oxy]methyl}tetrahydro-2*H*-pyran-2-yl imidoformate and *Amino(1,8-naphthyridin-3-yloxy) acetic acid are important for strong hydrogen bonding interaction with the inhibitors. Based on these findings, we have developed the hypothesis that the inhibition of Bcl2L10 is possible by some drugs.

References

1. Aouacheria A, Arnaud E, Venet S, Lalle P, Gouy M, Rigal D and Gillet G (2001) NrH, a human homologue of Nr-13 associates with Bcl-Xs and is an inhibitor of apoptosis. Oncogene. 20(41):5846-5855.

2. Bates S, Phillips AC, Clark PA, Stott F, Peters G, Ludwig RL, Vousden KH. (1998) p14ARF links the tumour suppressors RB and p53. Nature 395:124-125

3. Bell S, Klein C, Muller L, Hansen S, Buchner J. (2002). p53 contains large unstructured regions in its native state. J Mol Biol, 322:917-927

4. Bischoff JR, Kirn DH, Williams A, Heise C, Horn S, Muna M, Ng L, Nye JA, Sampson-Johannes A, Fattaey A, McCormick F. (1996). An adenovirus mutant that replicates selectively in p53-deficient human tumor cells. Science, 274:373-376

5. Biesalski HK, Bueno de Mesquita B, Chesson A, et al. (1998). "European Consensus Statement on Lung Cancer: risk factors and prevention. Lung Cancer Panel". CA: a cancer journal for clinicians 48 (3): 167–76.

6. Blagosklonny, MV. (2002). P53: An ubiquitous target of anticancer drugs. International Journal of Cancer, 98:161-166

7. Brune, B. et al (1999) "Nitric oxide (NO): an effector of apoptosisname". Cell Death and Differentiation. 6(10): 969-975.

8. Dejean L M, Martinez-Caballero S and Kinnally K W (2006) Is MAC the knife that cuts cytochrome c from mitochondria during apoptosis? . Cell Death and Differentiation. 13(8):1387–1395.

9. Ewing T J A and Kuntz I D (1996) Critical Evaluation of Search Algorithms for Automated Molecular Docking and Database Screening. J. of Computational Chem. 18(9):1175-1189.

10. Friesner R A, Banks J L, Murphy RB, Halgren T A, Klicic J J, Mainz D T, Repasky M P, Knoll E H, Shelley M, Perry J K, Shaw D E, Francis P and Shenkin P S (2004) Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. J. Med. Chem. 47 (7):1739–1749.

11. Hockenbery D, Nunez G, Milliman C, Schreiber R D and Korsmeyer S J (1990) Bcl-2 is an inner mitochondrial membrane protein that blocks programmed cell death. Nature. 348:334.

12. Huynen M, Dandekar T and Bork P (1998) Differential genome analysis applied to the species-specific features of *Helicobacter pylori*. FEBS Lett. 426(1):1-5.

13. Jemal, A; Bray, F, Center, MM, Ferlay, J, Ward, E, Forman, D (2011-02-04). "Global cancer statistics". CA: a cancer journal for clinicians 61 (2): 69–90.

14. O'Dell, edited by Michael D. Stubblefield, Michael W. (2009). Cancer rehabilitation principles and practice. New York: Demos Medical. pp. 983. ISBN 978-1-933864-33-4.

15. Julia Kravchenko; Igor Akushevich; Manton, Kenneth G. (2009). Cancer mortality and morbidity patterns in the U. S. population: an interdisciplinary approach. Berlin: Springer. p. 118. ISBN 0-387-78192-7.

16. Kang M H and Reynolds C P (2009) Bcl-2 Inhibitors: Targeting Mitochondrial Apoptotic Pathways in Cancer Therapy. Clin Cancer Res -Molecular pathway. 15:1126-1132.

17. Ke N, Godzik A and Reed J C J (2001) Bcl-B, a novel Bcl-2 family member that differentially binds and regulates Bax and Bak. J Biol Chem. 276(16):12481-12484.

18. Kearsley S K, Underwood D J, Sheridan R P and Miller M D (1994). Flexibases: a way to enhance the use of molecular docking methods. J. Comput. Aided Mol. Des. 8(5): 565–582.

19. Kihlmark M, Imreh G and Hallberg E (2001) Sequential degradation of proteins from the nuclear envelope during apoptosis. Journal of Cell Science 114(20):3643–53.

20. Klebe G and Mietzner T (1994) A fast and efficient method to generate biologically relevant conformations. J. Comput. Aided Mol. Des. 8(5): 583–606.

21.Krajewski S, Tanaka S, Takayama S, Schibler MJ, Fenton W and Reed JC (1993) Investigation of the subcellular distribution of the Bcl- 2 oncoprotein: Residence in the nuclear envelope, endoplasmic reticulum, and outer mitochondrial membranes. Cancer Res. 53(19):470l-4714.

22. Kuper H, Adami HO, Boffetta P (June 2002). "Tobacco use, cancer causation and public health impact". Journal of internal medicine 251 (6): 455–66.

23. Letai A (2003) BH3 domains as BCL-2 inhibitors: prototype cancer therapeutics. Expert Opin Biol Ther. 3(2):293-304.

24. McCormick F. (2001). Cancer gene therapy: fringe or cutting edge? Nat Rev Cancer, 1:130-141.

25. Monaghan P. Robertson D, Amos T A S, Dyer M J S, Mason D Y and Greaves M F (1992) Ultrastructural localization of Bcl-2 protein. J Histochem Cytochem 40:I819-1825.

26. Nagata S (2000) Apoptotic DNA fragmentation. Exp. Cell Res. 256 (1): 12–18.

27. Oltersdorf T, Elmore S W, Shoemaker A R, Armstrong R C, Augeri D J, Belli B A, Bruncko M, Deckwerth T L, Dinges J, Hajduk P J, Joseph M K, Kitada S, Korsmeyer S J, Kunzer A R, Letai A, Li C, Mitten M J, David G. Nettesheim, Ng S C, Nimmer P M, O'Connor J M, Oleksijew A, Petros A M, Reed J C, Shen W, Tahir S K, Thompson C B, Tomaselli K J, Wang B, Wendt M D, Zhang H, Fesik S W and Rosenberg S H (2008) An inhibitor of Bcl-2 family proteins induces regression of solid tumours. Nature Reviews- Molecular Cell Biology. **9**:47-59.

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

29. Rautureau G J, Day C L and Hinds M G (2010)The structure of Boo/Diva reveals a divergent Bcl-2 protein. Proteins. 78(9):2181-2186.

30. Sasco AJ, Secretan MB, Straif K (August 2004). "Tobacco smoking and cancer: a brief review of recent epidemiological evidence". Lung cancer (Amsterdam, Netherlands) 45 Suppl 2: S3–9.

31. Schütze, M; Boeing, H, Pischon, T, Rehm, J, Kehoe, T, Gmel, G, Olsen, A, Tjønneland, AM, Dahm, CC, Overvad, K, Clavel-Chapelon, F, Boutron-Ruault, MC, Trichopoulou, A, Benetou, V, Zylis, D, Kaaks, R, Rohrmann, S, Palli, D, Berrino, F, Tumino, R, Vineis, P, Rodríguez, L, Agudo, A, Sánchez, MJ, Dorronsoro, M, Chirlaque, MD, Barricarte, A, Peeters, PH, van Gils, CH, Khaw, KT, Wareham, N, Allen, NE, Key, TJ, Boffetta, P, Slimani, N, Jenab, M, Romaguera, D, Wark, PA, Riboli, E, Bergmann, MM (2011-04-07). "Alcohol attributable burden of incidence of cancer in eight European countries based on results from prospective cohort study". BMJ (Clinical research ed.) 342: 20-32.

32. Seitz HK, Pöschl G, Simanowski UA (1998). "Alcohol and cancer". Recent developments in alcoholism : an official publication of the American Medical Society on Alcoholism, the Research Society on Alcoholism, and the National Council on Alcoholism 14: 67–95.

33. Suresh P S, Kumar A, Kumar R and Singh V P (2008) An in silico approach to bioremediation: laccase as a case study. J. Mol. Graph. Model. 26(5):845–849.

34. Susin S A, Daugas E, Ravagnan L, Samejima K, Zamzami N, Loeffler M, Costantini P and Ferri KF et al. (2000) Two Distinct Pathways Leading to Nuclear Apoptosis. Journal of Experimental Medicine 192(4):571–580.

35. Vaux D L, Cory S and Adam J M (1988) Bcl-2 gene promotes haemopoietic cell survival and cooperates with c-myc to immortalize pre- B cells. Nature. 335(29):440-443.

36. Vogler M, D Dinsdale, M J S Dyer and G M Cohen (2008) Bcl-2 inhibitors: small molecules with a big impact on cancer therapy. Cell Death Differ. 16(3):360-367.

37. Vogelstein B, Lane D, Levine AJ. (2000). Surfing the p53 network. Nature, 408:307-310 Anand P, Kunnumakkara AB, Kunnumakara AB, et al. (September 2008). "Cancer is a Preventable Disease that Requires Major Lifestyle Changes". Pharm. Res. 25 (9): 2097–116.

38. Wang Q and Pang Y P (2007) Romesberg, Floyd. ed. Preference of small molecules for local minimum conformations when binding to proteins. PLoS ONE 2(9):820(1-10).

39. Youle R J and Strasser A (2009) The BCL-2 protein family: opposing activities that mediate cell death. Cell Death and Differentiation. 16:360–367.

40. Zhang H, Holzgreve W and De Geyter C (2001) Bcl2-L-10, a novel anti-apoptotic member of the Bcl-2 family, blocks apoptosis in the mitochondria death pathway but not in the death receptor pathway. Hum. Mol. Genet. 10:2329-2339.

41. Zody M C, Garber M, Sharpe T, Young S K, Rowen L, O'Neill K., Whittaker C A, Kamal M , Chang J L, Cuomo C A, Dewar K, FitzGerald M G, Kodira C D, Madan A, Qin S, Yang X, Abbasi N, Abouelleil A and Nusbaum C (2006) Nature 440:671-675.

42 . Zsoldos Z, Reid D, Simon A, Sadjad S B and Johnson A P (2007) eHiTS: A new fast, exhaustive flexible ligand docking system. Journal of Molecular Graphics and Modelling. 26(1):198–212.